

PROTEOLYSIS REGULATORS IN *DROSOPHILA MELANOGASTER* SEMINAL FLUID

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Brooke Ashley Laflamme

August 2012

© 2012 Brooke Ashley Laflamme



# STUDIES OF PROTEOLYSIS REGULATORS IN *DROSOPHILA MELANOGASTER* SEMINAL FLUID

Brooke Ashley Laflamme, Ph. D.

Cornell University 2012

In *Drosophila melanogaster*, seminal fluid proteins (Sfps) are transferred to females during mating and, together with sperm, are necessary for the many post-mating responses elicited in females. Identifying the functions of all individual Sfps is challenging, given the large number of Sfps and the possibility that many have redundant functions. One way to tackle this challenge is to focus on key regulators of pathways that involve multiple Sfps. Proteases are good candidates for these regulators. Proteases and protease inhibitors have been identified in the ejaculates of animals in taxa ranging from invertebrates to mammals, and form a major protein class among *Drosophila* Sfps. Other than a single protease cascade in mammals that regulates seminal clot liquefaction, no proteolytic cascades (*i.e.* pathways with at least two proteases acting in sequence) have been identified in seminal fluids. Though several *Drosophila* Sfps are proteolytically cleaved either during or after mating, little is known about the proteases involved in these cleavage events or the physiological consequences to the female of proteolytic activity in the seminal fluid.

This thesis addresses the functions of Sfp proteases in *Drosophila*. Chapter 1 provides an overview of seminal fluid proteolysis regulators. In Chapter 2, I report that a *D. melanogaster* Sfp, the predicted serine protease ‘seminase’, coordinates two pathways in the mated female.

One leads to the proper localization and long-term function of the Sfp sex peptide. The second leads to activation of the predicted astacin metalloprotease CG11864, through an apparent protease cascade. Chapter 3 presents experimental evidence that CG11864 is activated by proteolysis at a predicted pro-peptide cleavage site and that the male and female coordinate for efficient processing of CG11864's substrates. In Chapter 4, I focus on the proteolytic processing of ovulin and describe methods to quantify its processing rate that can be used in the future to identify additional proteins required for ovulin processing. Chapter 5 describes a computational method used to identify candidate Sfps, including proteolysis regulators, based on transcriptional profiles, and provides experimental validation of those predictions. Finally, Chapter 6 discusses future directions for research of the seminase/CG11864 protease cascade.

## BIOGRAPHICAL SKETCH

Brooke Ashley LaFlamme was born January 25, 1984 in Tucson, Arizona. She has two sisters, Tempest and Brandi, and four nieces and nephews: Sonya, Jordan, Michael, and Madison. Brooke attended Amphitheater High School in Tucson from 1998-2002. In 2002, with a scholarship in hand, Brooke became the first in her family to attend a four-year university. In her freshman year at the University of Arizona, Brooke took a general education science class with Prof. Harold P. Larson which changed the course of her life: she had never before considered becoming a scientist, but due to Prof. Larson's inspiring lectures on the power of scientific discovery, she changed her major from English to a double major in Molecular & Cellular Biology and Ecology & Evolutionary Biology (with minors in chemistry and Japanese).

While at the UofA, Brooke became involved in research in the lab of Prof. Therese Markow under the direct supervision of then Ph.D. student Laura Reed. In Prof. Markow's lab, Brooke studied the genetic architecture of male sterility in hybrids of two desert *Drosophila* species and later worked on a project with Prof. Nancy Moran identifying endosymbiotic bacteria in the ovaries of several *Drosophila* species. As part of the second project, Brooke was able to travel to Japan during the summer of 2005 to do research in the lab of Dr. Takema Fukatsu (and practice her Japanese skills). During her three years in the Markow lab, Brooke learned about the challenges and rewards of scientific research and decided to pursue her Ph.D. in genetics after graduating from the UofA in 2006.

Having lived her life up to that point in the Sonoran Desert, Brooke decided she needed a change of scenery. She moved to Ithaca, NY and entered the graduate field of Genetics and Development at Cornell University. She conducted her thesis research in the lab of Prof. Mariana Wolfner. During those six years, she learned how to weather the cold winters and run countless

Western blots. She also met the man she would later marry. A fellow graduate student introduced Brooke to a very nice Hungarian physics Ph.D. student, Attila János Bergou, in the spring of 2007. On July 13, 2013, Brooke and Attila will be wed on the shore of Lake Balaton in Hungary. Upon leaving Cornell, Brooke will pursue a postdoctoral position in New York City.

For Tempest and Brandi.  
Here's to beating the odds.

## ACKNOWLEDGMENTS

The work in this thesis could not have been accomplished without the contributions of many people. I first, and foremost, want to thank my absolutely wonderful advisor, Mariana Wolfner, for her guidance, support, and encouragement and for providing the best lab environment for my Ph.D. work that I could have imagined. Mariana has given me countless pep talks during those times that grad school and challenges with my research seemed too daunting, and I am very grateful for that. I also want to thank the other members of my committee, Andrew Clark, Dan Barbash, and Jason Mezey, for their helpful comments and discussions over these six years.

No research is ever done completely alone, and the work in my thesis is no exception. Nearly all of the research presented here is the result of collaborations with members of the Wolfner lab and others. I especially need to thank K. Ravi Ram, Frank Avila, Laura Sirot, Julien Ayroles, and Margarida Moreira for their direct contributions to my thesis work. I am also grateful to the other members of the Wolfner lab, past and present, for their comments and suggestions, and for simply being wonderful lab members: Alex Wong, Kate Sackton, Vanessa Horner, Jun Cui, Jessica Sitnik, Dustin Rubinstein, Clement Chow, Jennifer Apger, Geoff Findlay, Amber Krauchunas, Caroline Sartain, Dorina Frasheri, Jamie Moore, and above all, Norene Buehner.

On a more personal note, I want to acknowledge the emotional contributions to the completion of my graduate work made by my friends, which are too numerous to list here. Without them, the ups and downs of graduate school would have seemed like much more down than up. I want to thank my fiancé, Dr. Attila Bergou, and his family for their endless love, support, and encouragement. I also want to thank my sister, Tempest LaFlamme Spears, for being proud of me and blindly supportive of my decision to pursue a Ph.D.

Finally, I want to acknowledge funding from the National Science Foundation's Graduate Research Fellowship Program and from the Cornell University Presidential Life Sciences Fellowship.

## TABLE OF CONTENTS

<u>Section Name</u>	<u>Page Number</u>
BIOGRAPHICAL SKETCH	iii
DEDICATION	v
ACKNOWLEDGEMENTS	vi
CHAPTER 1: Overview of proteolysis regulators in the seminal fluid	1
CHAPTER 2: The <i>Drosophila melanogaster</i> seminal fluid protease ‘seminase’ regulates proteolytic and post-mating reproductive processes	37
CHAPTER 3: Mechanistic studies of the CG11864-mediated proteolysis of <i>Drosophila melanogaster</i> seminal fluid proteins	65
CHAPTER 4: Analysis of ovulin processing rate variation in inbred lines of <i>Drosophila melanogaster</i>	109
CHAPTER 5: Functional genome annotation of <i>Drosophila</i> seminal fluid proteins using transcriptional genetic networks	143
CHAPTER 6: Discussion and future directions	160
APPENDIX 1: <i>In vitro</i> expression of ovulin and proteolytic enzymes	169
APPENDIX 2: Molecular social interactions: <i>Drosophila melanogaster</i> seminal fluid proteins as a case study	181
APPENDIX 3: Investigating the role of mating in expression of <i>Drosophila</i> oogenesis-related genes	189
REFERENCES	202

## LIST OF FIGURES

<u>Figure Number and Title</u>	<u>Page Number</u>
Figure 2.1: Processing of SFPs is defective in the absence of seminase	43
Figure 2.2: Levels of seminase protein are reduced in RNAi lines	44
Figure 2.3: Seminase is produced in the accessory glands and is processed during and after mating	46
Figure 2.4: Females mated to seminase knockdown males lay fewer eggs	48
Figure 2.5: Egg-laying, hatchability, and sperm storage in females mated to seminase Line 2 males.	49
Figure 2.6: Females mated to seminase knockdown males retain more sperm 4 and 10 days after mating	51
Figure 2.7: LTR proteins fail to accumulate in the seminal receptacle in the absence of seminase	53
Figure 2.8: Seminase is required for two post-mating pathways	58
Figure 3.1: Schematic of ovulin cleavage pattern	71
Figure 3.2: DGRP-517 males' seminal fluid lacks ovulin-cleaving activity linked to CG11864	72
Figure 3.3: Genomic location of blood element insertion in DGRP-517	75
Figure 3.4: Backcrossed CG11864 null males retain processing defects	77
Figure 3.5: UAS-CG11864 expression in CG11864 null males and in virgin females	80
Figure 3.6: CG11864 expressed in females can process ovulin	83



<u>Figure Number and Title</u>	<u>Page Number</u>
Figure 3.7: Ovulin can be partially cleaved in ejaculates of males in the absence of females	87
Figure 3.8: Predicted activational cleavage site of CG11864 is required for ovulin and Acp36DE processing	91
Figure 3.9: Conserved catalytic site mutation in CG11864 results in slower ovulin processing	94
Figure 4.1: Ovulin cleavage products and rate categories	114
Figure 4.2: Ovulin cleavage sites determined by Laura Sirot	118
Figure 4.3: Quantification of ovulin processing rate as measured using Western blots for 12 inbred lines	121
Figure 4.4: DGRP lines with fast or slow ovulin processing	124
Figure 4.5: Ovulin CP3 is detected in mated females by SRM mass spec	129
Figure 4.6: Ovulin CP3 is detected at highest levels in WT matings at 1 hour ASM	131
Figure 4.7: Orbitrap data for DGRP-786 vs. DGRP-852 pilot experiment	134
Figure 4.8: Qtrap data for DGRP-786 vs. DGRP-852 pilot experiment	136
Figure 4.9: Blots used for categorical ovulin processing rate study	139
Figure 5.1: Graphical representation of the correlation of among known SFPs	151
Figure 5.2: Gene expression level correlation between cSFPs and CG9997	156
Figure 5.3: RT-PCR analysis of candidate gene expression	157

<u>Figure Number and Title</u>	<u>Page Number</u>
Figure A1.1: Ovulin and CG11864 are expressed in vitro in an E. coli extract system	170
Figure A1.2: Ovulin and CG11864 expression with eukaryotic in vitro expression systems	172
Figure A1.3: Semi in vitro ovulin processing assays	173
Figure A1.4: Proteins expressed in insect cell culture using the baculovirus expression system	176
Figure A1.5: Semi in vitro ovulin processing assay with Sf9-expressed proteins	177
Figure A1.6: Dialysis and concentration of Sf9-expressed proteins	178
Figure A1.7: N-terminal mapping of Sf9-expressed CG11864 and “active” CG11864	179
Figure A2.1: Male and female genotypes affect outcome of sperm competition experiments	185
Figure A2.2: Sexual conflict may maintain polymorphisms in the population	186
Figure A3.1: Standard curves from qRT-PCR assays of oogenesis genes	193
Figure A3.2: qRT-PCR results separated by oocyte stage and mating status	195
Figure A3.3: Representative plots of chorion protein gene expression, separated by oocytes stage and mating status	197
Figure A3.4: Representative plots of ecdysone-responsive gene expression, separated by oocytes stage and mating status	199

## LIST OF TABLES

<u>Table Number and Title</u>	<u>Page Number</u>
Table 1.1: Human seminal fluid proteases	5
Table 1.2: Human seminal fluid protease inhibitors	8
Table 1.3: Proteases, inhibitors, and protease homologs in <i>Drosophila melanogaster</i> seminal fluid	11
Table 2.1: Female receptivity to second mating	50
Table 3.1: DGRP lines screened for ovulin processing defects	70
Table 3.2: Fertility assays performed with CG11864 null males	95
Table 4.1: Ovulin processing rate variation in 25 inbred lines	125
Table 4.2: Ovulin peptides from females mated to WT (CS) or 517 males	130
Table 4.3: Ovulin peptides from DGRP lines 852 and 786	137
Table 5.1: Genes selected for experimental validation	153
Table A1.1: List of constructs mentioned in Appendix 1	180
Table A3.1: Primers used for qRT-PCR of oogenesis genes	191
Table A3.2: Genes used for oogenesis study	192
Table A3.3: Results from standard curves	194

## CHAPTER 1

### OVERVIEW OF PROTEOLYSIS REGULATORS IN THE SEMINAL FLUID

#### INTRODUCTION

Seminal fluid—the non-sperm components of the male ejaculate—contains hundreds of proteins as well as non-protein components that, together, are vital for male fertility [1-3]. The functions of seminal fluid proteins (Sfps) in insects and mammals have been recently reviewed [3]. In aggregate, Sfps are vital to male fertility, playing important roles in sperm activation and storage [4-13]; pregnancy establishment [4, 14-18]; female immune function [15, 16, 19]; ovulation (in stimulated ovulators) [20-23]; and female post-mating behaviors [24, 25]. Individually, however, only a small proportion of Sfps have been assigned functions, and the molecular mechanisms for their functions are even less well understood.

Individual Sfp homologs are difficult to identify between distantly related organisms as a result of the rapid evolution of many male reproductive proteins [26-30]. On the other hand, the major protein classes that comprise Sfps are conserved between taxa as distant as insects and mammals [31]. Proteases and their inhibitors are among these major conserved classes. The function of specific male-derived proteases in the seminal fluid will be the major focus of this thesis.

Due to the complexity of the seminal fluid, identifying functions for individual seminal fluid proteins (Sfps) has been very challenging. One way to tackle this complexity is to focus on proteins that are likely to coordinate the functions of multiple other Sfps. Proteolysis regulators (proteases and their inhibitors) are one such protein class in the seminal fluid because they have

the ability to regulate functions of multiple proteins through degradation, removal of inhibitory peptides, or through processing of prohormones. A focus on the substrates and fertility consequences of proteolysis regulators in the seminal fluid may lead to a more integrated, pathway-centric understanding of Sfp function.

Proteolysis is central to many vital biological processes, such as immunity, blood clotting, cell cycle regulation, and tissue morphogenesis [32, 33]. An external signal can be rapidly amplified through the activation of a single protease that regulates multiple downstream pathways [34, 35]. Serine proteases in particular can also be rapidly, and irreversibly, inactivated through the action of protease inhibitors [34-37]. Both activation and inhibition of proteases can be achieved without new protein synthesis, making them ideal for regulating extracellular processes such as interactions between male and female proteins following ejaculation in internally fertilizing animals. Despite this, the roles of seminal fluid proteases in fertility remain poorly understood.

### ***Proteolysis regulators are common constituents of seminal fluid***

Proteomic and transcriptomic studies of the seminal fluid of many animals, including humans, have identified large numbers of Sfp proteolysis regulators. The most thorough proteomic studies are those from human and *Drosophila* seminal fluid, which I will focus on in this chapter. Sfp proteases in humans and insects fall into several protease classes, including metalloproteases, cysteine proteases, and aspartic proteases, though the majority are trypsin- or chymotrypsin-like serine proteases (Tables 1.1-1.3).

Proteases are classified based on their hydrolysis mechanism [38]: serine proteases have a conserved catalytic triad consisting of a His, Ser, and Asp that coordinates a water molecule.

The serine residue acts as a nucleophile to attack the carbonyl carbon of the substrate's scissile peptide bond [38]. Serine proteases are the most common protease class in both insects and mammals [33, 39-41]. Metalloproteases are so named because they use a metal ion (such as zinc) to polarize a water molecule within the active site; the water molecule is then used to hydrolyze the scissile peptide bond of the substrate [38]. The extracellular matrix metalloproteases (MMPs) [42] and the astacin metalloproteases [43] are important members of this class. Cysteine proteases use a nucleophilic cysteine for hydrolysis [38]. Cysteine proteases are most common in plants [44] but they are also very important in human physiology [45]. For example, many cathepsins [46], which are present in seminal fluid, are cysteine proteases. Finally, aspartic proteases, the first class of proteases to be described, use an aspartate as their catalytic residue [38]. The digestive proteases pepsin and gastricin (which is also a constituent of the seminal fluid [47, 48]) are examples of aspartic proteases [49].

Proteolysis must be tightly regulated to prevent premature activation of pathways or tissue damage that may result from overactive proteases. Protease inhibitors play an important part in regulation of proteolytic activity. As serine proteases are the most common class of proteases in seminal fluid, serine protease inhibitors are also the most prevalent class of protease inhibitors, though cysteine protease inhibitors (e.g., cathepsins) are also common. The prevalence of serine proteases and their inhibitors in the seminal fluid is expected, given the large proportion of these classes in the proteome (of both *Drosophila* and humans) [41, 50]. In addition to catalytically active proteases and protease inhibitors, two related classes of proteins are represented in *Drosophila* seminal fluid: catalytically inactive protease homologs and non-inhibitory serpins.

### ***Proteolysis regulators in human seminal fluid***

A large number of proteases (Table 1.1) and protease inhibitors (Table 1.2) have been identified in human seminal plasma. The majority of these proteins are known from empirical studies to act as either proteases or protease inhibitors, though a small number are classified as proteolysis regulators based on computational prediction. Ejaculate samples can be easily obtained from many mammalian species, including humans, to allow direct proteomic analysis of the seminal fluid, with or without the presence of sperm [reviewed in 51]. Fung, *et al.* [48] used 1D and 2D PAGE followed by mass spectrometry to identify over 100 proteins in normal human seminal fluid, including the serine protease prostate-specific antigen (PSA) and its major substrates, Semenogelins I and II. Many of these proteins, PSA included, were present in multiple modified forms, suggesting a role for post-translational modification in their regulation. In a later study, Pilch, *et al.* [52] took advantage of recent advances in proteomics to create a high-confidence list of 923 human seminal plasma proteins. Based on current protein annotations, this study identified 58 non-proteasomal proteases and peptidases, and 38 protease inhibitors. A complementary proteomic analysis of human prostasomes identified a total of 139 proteins, which included 1 protease not identified in the seminal fluid studies, the glutamate carboxypeptidase CPGL [47]; the other proteases identified in this study overlapped with the other proteomic studies of the seminal plasma [48, 52].

**Table 1.1: Human seminal fluid proteases.**

Proteases identified in human seminal plasma or prostasomes, listed by protease class. Classes are based either on experimental data or computational prediction. GI: NCBI identifier.



Protease class	Protein name	GI	References
<b>ADAM metalloprotease</b>	ADAMTS-1	1264483	[52]
	ADAM 7	296439449	[52]
<b>Aminopeptidase</b>	LAP3	48146373	[52]
	Aminoacylproline aminopeptidase	68566146	[52]
	Aminopeptidase N	143811362	[47, 52]
<b>Asparaginyl peptidase</b>	Legumain	56682964	[52]
<b>Aspartate protease</b>	Gastricin	129796	[47, 48]
<b>Carboxypeptidase</b>	Carboxypeptidase C	11055992	[52]
	Carboxypeptidase D	25089854	[52]
	Carboxypeptidase E	4503009	[52]
	Carboxypeptidase M	53832021	[52]
	Carboxypeptidase O	27436871	[52]
	Carboxypeptidase Z	62388875	[52]
	Glutamate carboxypeptidase-like protein 1	23396498	[52]
	Pancreatic carboxypeptidase B1	54607080	[52]
	Probable serine carboxypeptidase CPVL precursor	67476930	[52]
	Prolylcarboxypeptidase	4826940	[52]
	Glutamate carboxypeptidase (CPGL)	15620780	[47]
<b>Caspase</b>	Caspase 14	6912286	[52]
	Paracaspase	20455075	[52]
<b>Papain-like cysteine protease</b>	Cathepsin B	22538437	[52]
	Cathepsin D	4503143	[52]
	Cathepsin F	6042196	[52]
	Cathepsin G	4503149	[52]
	Cathepsin H	21264388	[52]
	Cathepsin O	4557501	[52]
	Cathepsin L1	22202619	[52]
	Cathepsin Z	22538442	[52]
	PPGB	56001072	[52]
<b>Other cysteine protease</b>	Calpain 1	12408656	[52]

**Table 1.1 (continued)**

<b>Protease class</b>	<b>Protein name</b>	<b>GI</b>	<b>Reference</b>
<b>Matrix metalloprotease</b>	MMP2	11342666	[52]
	MMP7	4505219	[52]
	MMP14	317373419	[52]
<b>Other Peptidase</b>	Dipeptidase 3	193211608	[52]
	Dipeptidyl-peptidase II	21617867	[52]
	Dipeptidyl-peptidase III	20532389	[52]
	Tripeptidyl-peptidase I	62896635	[52]
	Tripeptidyl peptidase II	55661754	[52]
	Peptidase A	23396498	[52]
	Puromycin-sensitive aminopeptidase	51704228	[52]
	Insulin protease	215274252	[52]
	Dipeptidyl peptidase IV	1352311	[47, 52]
	Neprilysin	128062	[47]
<b>Serine protease</b>	KLK12	37182171	[53]
	KLK14	156257445	[53]
	KLK15	9957760	[53]
	KLK4	64653115	[53]
	KLK5	37183138	[53]
	KLK8	37183190	[53]
	KLK11	5803199	[52, 53]
	KLK2	5031829	[52, 53]
	Acrosin light chain	115502349	[52]
	Brain-specific serine protease 4	11545839	[52]
	C1r-like serine protease analog protein	182705204	[52]
	complement component 1	62897163	[52]
	complement component 2	74755562	[52]
	complement factor B	67782358	[52]
	complement factor I	317373341	[52]
	matriptase	13124575	[52]
	prostasin	4506153	[52]
	serine protease 1	4506145	[52]
	TMPRSS2	115502469	[47, 52]
	PSA (KLK3)	4502173	[47, 48, 52]
<b>Thermolysin-like protease</b>	Angiotensin I converting enzyme	4503273	[52]
	Angiotensin-converting enzyme 2	48525714	[52]
<b>Ubiquitin processing protease</b>	Deubiquitinating enzyme 14	1729927	[52]

**Table 1.2: Human seminal fluid protease inhibitors.**

Protease inhibitors identified in human seminal plasma, listed by protease inhibitor class. Protease classes are based either on experimental data or computational prediction. GI: NCBI identifier.

<b>Protease inhibitor class</b>	<b>Protein name</b>	<b>GI</b>	<b>Reference</b>
<b>Cysteine protease inhibitor</b>			
	Cystatin A	48145945	[52]
	Cystatin SN	118188	[52]
	CST6	74724686	[52]
	Cystatin C	4503107	[48]
	Cystatin S	4503109	[48]
<b>Metalloprotease inhibitor</b>			
	Tissue inhibitor of metalloproteinase 1	4507509	[52]
	Tissue inhibitor of metalloproteinase 2	4507511	[52]
	Tissue inhibitor of metalloproteinase 3	4507513	[52]
<b>Peptidase inhibitor</b>			
	Peptidase inhibitor 15	7705676	[52]
<b>Serine protease inhibitor</b>			
	Alpha-1-antitrypsin	1703025	[48, 52]
	Plasma serine protease inhibitor	194018472	[48]
	Acrosin-trypsin inhibitor	123985	[52]
	Alpha-1-antichymotrypsin	112874	[52]
	Alpha-2-antiplasmin	112907	[52]
	Amyloid-beta A4 protein	112927	[52]
	Amyloid-like protein 2	1703344	[52]
	Angiotensinogen	113880	[52]
	Anosmin-1	134048661	[52]
	Antileukoprotease I	113636	[52]
	Antithrombin III	113936	[52]
	CD109 (TGF-beta-1-binding protein)	117949389	[52]
	Epididymal secretory protein E4	20141958	[52]
	Inter-alpha-trypsin inhibitor heavy chain H5	187609608	[52]
	Kallistatin	68067608	[52]
<b>Serine protease inhibitor</b>			
	Kunitz-type protease inhibitor 1	61252335	[52]
	Kunitz-type protease inhibitor 3	152031689	[52]
	Kunitz-type protease inhibitor 4	74749362	[52]
	Maspin (serpin B5)	229462757	[52]
	Neuroserpin (serpin I1)	3183087	[52]
	PEBP-1	1352726	[52]
	PEDF (serpin F1)	313104314	[52]
	Peptidase inhibitor 15 (SugarCrisp)	74735410	[52]
	Placental thrombin inhibitor (serpin B6)	161784343	[52]
	Plasma protease C1 inhibitor	124096	[52]
	Protein AMBP	122801	[52]
	Protein C inhibitor (serpin A5)	322510122	[52]
	Putative protease inhibitor WAP8	126302615	[52]
	Squamous cell carcinoma antigen 1 (serpin B3)	20141712	[52]
	Thyroxine-binding globulin (serpin A7)	1351236	[52]
	Tissue factor pathway inhibitor 2	1351226	[52]
	Transcortin (serpin A6)	115851	[52]
	SPINT3	189571689	[54]
	SPINT4	110626171	[54]
	SPINT5 (sequence only available from mouse)	53794664	[54]

Together, these proteomic studies of human seminal fluid [48, 52] and purified prostasomes [47] identified over 100 proteases, peptidases, or protease inhibitors (Tables 1.1 and 1.2). Several classes of proteases and protease inhibitors were identified in these studies. Of the proteolysis regulators identified in these three studies, there were 14 serine proteases and 32 serine protease inhibitors (these numbers do not exactly match those given in the original papers, as protein annotations have been updated; the updated information is used in tables 1.1 and 1.2). It is likely that more proteases remain to be identified, as even the most comprehensive of these studies failed to find more than three kallikrein-like serine proteases (KLKs 2, 3, and 11) [52], though at least nine are secreted into the seminal fluid [53]. Additional proteins may not be detected using these proteomic methods if they are products of novel, unannotated genes. Recently, three novel Sfp genes with Kunitz-family protease inhibitor domains were identified as expressed in human epididymis, and are likely secreted into the seminal plasma [54]. These genes have evolved recently and are much smaller than other protease inhibitor genes, making their identification by standard gene model prediction methods difficult [54].

### ***Proteolysis regulators in insect seminal fluid***

Proteolysis regulators are also highly represented in the seminal fluid of *Drosophila* as evidenced by several proteomic and transcriptomic studies (Table 1.3) [26, 55-60]. Recent proteomic studies of the *D. melanogaster* male reproductive tract have found that proteolysis regulators are expressed in all tissues of the male reproductive tract: accessory glands, testes, seminal vesicles, ejaculatory duct, and ejaculatory bulb [59, 61].

**Table 1.3: Proteases, inhibitors, and protease homologs in *Drosophila melanogaster* seminal fluid.**

Proteases, protease inhibitors, and their homologs in *Drosophila melanogaster* seminal fluid, listed by protein class. Most protein functions listed are based on predictions made by detection of conserved protease or protease inhibitor domains.

“Non-AG expression?” This column denotes whether gene expression is detected outside of the male accessory glands, based on data in FlyAtlas [62].

\* Highest gene expression level is outside the male accessory gland.

\*\* Annotation as protease inhibitor is based on sequence similarity to known serine protease inhibitor genes, rather than on direct detection of conserved domains.

Class	Gene name	Flybase identifier	Non- AG expression?	References
Serine protease	<i>seminase</i>	CG10586	No	[39, 57, 63]
	--	CG10587	No	[39, 57, 63]
	--	CG11037	No	[39, 57, 59]
	--	CG4815	Yes	[39, 57, 63]
	--	CG10041	No	[39, 57]
	--	CG33462	No	[39, 64]
Serine protease homolog	--	CG11664	No	[31, 39, 55, 65, 66]
	--	CG12558	No	[39, 63]
	--	CG14061	No	[57, 59]
	--	CG9997	No	[31, 63, 65-69]
	<i>sphinx-1</i>	CG32383	No	[39]
	<i>sphinx-2</i>	CG32382	No	[39, 63, 70]
	--	CG17242	No	[39, 57]
	--	CG32833	No	[39, 57, 58]
	--	CG34129	--	[31, 39, 40, 55, 57, 65-67]
	--	(previously CG6069) CG34130 (previously CG6069)	--	[31, 39, 40, 55, 57, 59, 65-67]
Serpins	<i>Spn1</i>	CG9456	Yes	[50, 59, 71]
	<i>Spn42Dd</i>			
	<i>Spn2</i>	CG8137	No	[31, 50, 55, 57-59, 65, 67, 71]
	<i>Spn28F</i>			
	<i>Spn3</i>	CG9334	No	[31, 50, 55-59, 65, 67, 71]
	<i>Spn38F</i>			
	<i>Spn7</i> <i>Spn28B</i>	CG6717	No	[72]
Noninhibitory serpin	<i>Acp76A</i>	CG3801	No	[31, 55, 73]
	<i>Spn76A</i>			
	<i>Spn53F</i>	CG10956	No	[31, 55, 73]
	<i>Spn75F</i>	CG32203	Yes	[30, 50, 55, 57-59, 67, 71]
	<i>Spn77Bb</i>	CG6663	Yes	[50, 57-59, 71]
	<i>Spn77Bc</i>	CG6289	Yes	[31, 50, 55, 57, 59, 65, 67, 71, 74]
MEROPS family 11 (Kazal) protease inhibitor	--	CG31704	No	[50, 57, 58]
	<i>Sfp33A3</i>	CG42474	No	[50, 57, 58]

**Table 1.3 (continued)**

<b>Class</b>	<b>Gene name</b>	<b>Flybase identifier</b>	<b>Non- AG expression?</b>	<b>References</b>
<b>MEROPS family 12 (Kunitz) protease inhibitor</b>	--	CG31515	No	[50, 57]
	<i>Acp24A4</i>	CG31779	Yes*	[50, 57, 58, 75]
	<i>Sfp23F**</i>	CG42459	--	[57, 58]
	<i>Sfp24Ba</i>	CG42461	--	[50, 57, 58]
	<i>Sfp24Bb**</i>	CG42462	--	[57, 58]
	<i>Sfp24Bc</i>	CG42602	--	[50, 58]
	<i>Sfp33A1**</i>	CG42472	--	[57, 58]
	--	CG42537	--	[59, 65]
		(previously CG10284)		
<b>MEROPS family 18 protease inhibitor (cysteine-rich trypsin inhibitor-like)</b>	<i>Acp62F</i>	CG1262	No	[50, 55, 57, 65, 76-80]
	<i>Sfp53D</i>	CG42477	No	[57, 58]
<b>Metalloprotease</b>	--	CG11864	No	[31, 55, 63, 65, 66, 78, 81]
	--	CG6168	No	[31, 55, 65-67, 79]
<b>Cysteine protease</b>	<i>Swim</i>	CG3074	Yes	[30]
	--	CG4847	Yes	[30, 31, 56-59]
<b>Cystatin (cysteine protease inhibitor)</b>	<i>Cystatin-like</i>	CG8050	Yes	[57]
<b>Aminopeptidase</b>	--	CG10576	Yes	[59]
<b>gamma-glutamyl transpeptidase</b>	<i>Ggt-1</i>	CG6461	Yes	[56, 57]
<b>Peptidase homolog</b>	--	CG9806	No	[82]
	--	CG6071	Yes	[57]



Findlay, *et al.* [57] used heavy isotope labeling of *Drosophila melanogaster* male proteins to identify Sfps that were transferred to the female reproductive tract, while excluding female proteins from the analysis. Using this method combined with shotgun proteomics, they found that over 20% of Sfps that they identified were proteolysis regulators (15 proteases and 14 protease inhibitors predicted among 138 high-confidence proteins) [57]. Further analysis of these data identified additional Sfps that were not previously predicted based on the *D. melanogaster* genome assembly, including protease inhibitors [58].

To date, these and other studies [26, 55] have identified 21 proteases and peptidases, 17 protease inhibitors, 10 predicted inactive serine protease homologs, and 5 predicted non-inhibitory serpin homologs in the seminal fluid of *D. melanogaster* (Table 1.3). Future work may uncover new Sfp proteolysis regulators in *Drosophila*. Several predicted proteolysis regulators were identified in a genome-wide expression study that uncovered over 100 genes with expression patterns correlated to those of known Sfps [60] (this study will be further discussed in Chapter 5). This gene list likely contains previously unknown Sfp genes, including proteolysis regulators and their homologs, as well as genes important for Sfp gene regulation.

Proteomic analysis of male accessory glands from a distant relative of *D. melanogaster*, *D. mojavensis*, also identified many proteases and protease inhibitors that are likely components of the seminal fluid [83]. Six predicted proteases and 1 predicted protease inhibitor had an Sfp homolog in *D. melanogaster* [57]. An additional 16 predicted proteases and 8 predicted protease inhibitors were *D. mojavensis*-specific Sfps [83], indicating that proteolysis regulators have species-specific roles in seminal fluid. This is consistent with data demonstrating that Sfps in *D. melanogaster* and *D. mojavensis* are, as a class, more likely to be evolving under positive selection than are other Sfp protein classes [26, 30, 84].

Recent work in other arthropod species has identified proteases and/or protease inhibitors in the seminal fluid of mosquitoes [85], sandflies [86], medflies [87], honeybees [88], butterflies and moths [89-91], *Tribolium* [92], crickets [93], and ticks [94]. Vertebrate species other than humans in which proteolysis regulators have been identified in the seminal fluid include mouse [95, 96], bull [97, 98], stallion [99], fish [100, 101], and birds [102, 103].

## **FUNCTIONS OF SEMINAL FLUID PROTEASES AND THEIR INHIBITORS**

Given the recent explosion in Sfp identification across diverse taxa, we can hope to uncover common themes for Sfp function as well as species-specific Sfp roles. Studies on Sfp proteolysis regulators, especially in model organisms, will aid in understanding of pathways mediated by seminal fluid. Below, I will highlight functional studies of proteolysis regulators in the seminal fluid.

### ***Sfp-mediated proteolysis and its regulation in insect ovulation and egg-laying***

In most insects, mating—specifically receipt of Sfps—induces high levels of oogenesis, ovulation, and egg-laying (for review, see [2]). Serine protease Sfps have been implicated in several species in this mating-induced increase in egg production. A trypsin-like serine protease in the *Allonemobius socius* complex of crickets is necessary for full induction of egg-laying after mating [22]. This protease, ejac-sp, is one of the most abundant proteins in the cricket spermatophore (a discrete package of seminal fluid and sperm that is deposited by the male for female insemination). The abundance of ejac-sp has been shown to decline with male age, along with the male's ability to induce egg-laying in females [22]. RNAi knockdown of the ejac-sp

protein in male crickets led to reduced levels of the protein in the ejaculate which correlated with reduced numbers of eggs laid in females mated to these males. Ejac-sp is also a candidate speciation gene, as it shows a high level of divergence due to positive selection between two related cricket species [104].

In *Drosophila melanogaster*, a predicted trypsin-like serine protease in the seminal fluid, *seminase* (CG10586), is also required for normal induction of egg-laying in mated females [63]. Females mated to *seminase* RNAi knockdown males lay eggs in similar numbers to controls during the first 24 hours after mating, but fail to maintain a high level of egg-laying in the following days, resulting in a large decrease in the total number of eggs laid over a 10 day experiment. This reduction appears to be due to a defect in sperm release from the seminal receptacle, the major sperm storage organ in *D. melanogaster* females. The placement of *seminase* into this pathway, which is dependent on the Sfp sex peptide, is further discussed in Chapter 2.

*Seminase* also has a seemingly unrelated role in initiating a protease cascade that leads to processing of Sfps involved in earlier post-mating processes (further discussed in Chapter 2) [63]. *Seminase* itself is cleaved during mating, while still in the male reproductive tract, suggestive of activational pro-peptide removal, and may be responsible for quickly activating multiple post-mating processes mediated by Sfps. *Seminase* is required for the predicted activational processing of an astacin-family zinc metalloprotease in the seminal fluid, CG11864. The cleavage of CG11864 is initiated in the male reproductive tract, in transit to the female [81]. The role of *seminase* in this proteolytic pathway is also discussed further in Chapter 2.

After deposition into the female tract, CG11864 is needed for processing of downstream Sfps [81], including the ovulation-inducing prohormone *ovulin* [105]. *Ovulin* is cleaved in three

sequential steps from its N-terminal end to give four final cleavage products [106]. The trypsin protease inhibitor Acp62F (another Sfp) also affects ovulin processing; knock-out of Acp62F results in slower ovulin processing [80]. In contrast, knock-down of the female-expressed inactive serine protease homolog CG9897 (also discussed in a later section below), results in faster ovulin processing [107]. The involvement of both male and female proteins in this proteolytic pathway suggests that precise regulation of ovulin processing is important for its physiological function.

Ovulin processing may lead to an increase in ovulin's activity, as ectopic expression of full-length ovulin as well as two C-terminal cleavage products can induce ovulation upon ectopic expression in virgin females [108]. However, no functional role for ovulin's cleavage products has been tested in the context of mating. Despite the unclear role of CG11864 in male fertility, the semine/CG11864 pathway is an interesting *in vivo* model for studying the regulation of protease pathways in seminal fluid.

### ***Protease pathways in semen coagulation***

Upon ejaculation in humans, proteins secreted by the seminal vesicles, Semenogelins I and II (SgI and SgII), and fibronectin, are mixed with other components of the semen to cause coagulation of the ejaculate [109]. Semenogelin (Sg) proteins create a matrix that immobilizes a portion of the ejaculated sperm in a structure called the seminal clot [110]. This gel-like structure is degraded within approximately 20-30 minutes post-ejaculation by the action of kallikrein-like proteases (KLKs) in the prostatic fluid in a process known as semen liquefaction [110]. KLKs are trypsin- or chymotrypsin-like serine proteases encoded by 15 genes at the chromosomal locus 19q13.4 [111]. KLKs are required both for semen liquefaction and skin desquamation in humans

[for review, see 112]. Mutations in KLKs and defects in their regulation have been linked to various cancers, including prostate cancer [111].

Semen liquefaction is the end result of a seminal fluid KLK protease cascade. KLKs 2, 3, 4, 5, 8, 11, 12, 14, and 15 are secreted by the prostate [reviewed in 53] and act in a protease cascade initiated by the reduction in  $\text{Zn}^{2+}$  concentration that follows mixing of the prostatic fluids with other seminal fluids [reviewed in 112]. Seminal liquefaction is required to release spermatozoa trapped by the seminal clot, and defects in liquefaction may lead to fertility problems [113].

Pro-KLK3 (the inactive form of KLK3, also known as prostate-specific antigen, PSA) is ultimately activated by the KLK cascade and directly cleaves the Sg proteins. Activated PSA is subsequently inactivated, in part, by the serine protease inhibitor, protein C inhibitor (PCI) [114]. PCI also forms tertiary structures with the Sg proteins, further contributing to the inactivation of PSA.

Which KLKs are directly responsible for activating PSA is not entirely clear. Williams, *et al.* [115] created transgenic mouse lines expressing human PSA and KLK2 in the prostate; these are the only two KLKs without homologs in the mouse [116]. They found that PSA was only active in the prostate when co-expressed with KLK2, suggesting that KLK2 may be important for regulation of PSA activity.

A recent study identified KLK14 as a potential activator and subsequent inhibitor of PSA [117]. Addition of active KLK14 to ejaculate samples rapidly and transiently increased chymotrypsin-like activity against a synthetic peptide (which could be attributed to PSA), compared to samples treated with a KLK14-specific inhibitor [117]. Longer incubations showed a reversal of this effect, with greater chymotrypsin-like activity in samples treated with the

KLK14 inhibitor relative to those treated with active KLK14. Samples treated with KLK14 also showed greater fragmentation of PSA, suggesting an inhibitory mechanism by internal cleavage. KLK14 may also directly degrade purified SgI and SgII *in vitro*, complementing the function of PSA [117]. Consistent with this, Emami, *et al.* found that patients with clinically delayed seminal clot liquefaction also had significantly lower KLK14 expression than normal patients [117].

Other seminal fluid serine proteases may also be involved in the KLK cascade. The serine protease plasmin (the activated form of plasminogen) is able to activate KLK11 *in vitro* [118]. In addition to proteins mentioned above, many of which are also expressed as blood plasma proteins, classic hemostatic factors may also contribute to seminal clotting and liquefaction [119-122]. These include a major role for tissue factor (TF; thromboplastin) [reviewed in 122]. Consistent with an important role of TF in seminal fluid, levels of this protein are much higher in semen (average of 21 ng/mL [123]) than in blood (average of 85 pg/mL [119]).

The extent to which the human seminal clotting/liquefaction pathways can be extrapolated to other species is unclear. Among primates, the presence and strength of selection on Sg genes appears to correlate with inferred levels of sperm competition (i.e.: the level of polyandry) [28, 124, 125]. The semen coagulum may help prevent fertilization by subsequent mates within a short time period after the first mating, it may prevent loss of sperm, or may promote fertilization by some as yet unknown mechanism, all of which could contribute to an advantage for one male over his rivals in sperm competition.

In chimpanzees, which experience high levels of sperm competition, the SgI protein has become greatly expanded [126], providing potentially more sites for cross-linking or non-covalent interactions during copulatory plug formation.

Sg genes show signatures of strong positive selection in chimpanzee [124, 127], as does transglutaminase 4 (TGM4), which functions in cross-linking Sg proteins in many species [28]. TGM4 also shows signatures of positive selection in bonobos [125]. In contrast, SgI, SgII, and TGM4 are all pseudogenes in gorilla, which experiences low, if any, sperm competition [28, 125-127]. Human SgII and TGM4 appear to be evolving nearly neutrally, though they are clearly functional, as opposed to the non-functional, pseudogenized SgI, SgII, and TGM4 genes in gorilla [125].

While TGM4 is not itself a protease, its function may be important in regulating proteolysis in the semen coagulum or in mating plugs of other species (mating plugs are similar, though not equivalent, structures to the human semen coagulum). For example, the mouse mating plug is formed by the crosslinking action of transglutaminase in the seminal fluid. Cross-linking of mating plug proteins makes the plug resistant to proteolysis, presumably by female-expressed proteases [128]. The co-evolutionary pattern of transglutaminase and its substrates in the seminal fluid [128] in crosslinking mouse mating plug proteins may be a response to the action of female proteases.

The mating plug may act to prevent insemination by subsequent males (see section on sperm competition below). In *Drosophila*, genetic deletion of the major mating plug protein, PEBII resulted in an increase in remating by females very soon after mating, relative to females mated first to normal males [129], though the mechanism for this function of PEBII is not known. In the mouse, unchecked proteolysis of the mating plug, due to the deletion of the seminal vesicle-expressed serine protease inhibitor, SERPINE2 (also known as protease nexin-1) in male mice led to malformed mating plugs and decreased fertility [130]. Plugs formed after mating with SERPINE2 mutant males were abnormally small, soft, and fibrous compared to

plugs formed by wildtype males, apparently due to proteolysis of important structural proteins of the mating plug. These defects resulted in loss of sperm from the female reproductive tract after mating. Even when sperm are able to remain in the female reproductive tract, the function of SERPINE2 as a decapacitation factor [10] likely contributes to the infertility seen in these mutant mice [10, 130].

Mating plugs in *Anopheles gambiae* promote sperm storage [131]. Analysis of the *Anopheles* mating plug revealed the presence of 5 male-derived proteases and many female-derived proteases [131]. A male accessory gland-specific transglutaminase is also required in *Anopheles* for semen coagulation. Crosslinking of the semen coagulum may prevent proteolysis by female proteases in mosquitoes, as in the mammalian examples above. Even though their mating systems are diverse, mating plugs in mammals and insects may play similar roles in male-female interactions.

### ***Roles for Sfp proteolysis and inhibitors in sperm storage and sperm competition***

Sperm storage is a universal phenomenon in animals with internal fertilization. The storage of sperm in specialized organs or areas within the female reproductive tract has three stages: entry of sperm into storage, maintenance of sperm within storage (for time periods between hours and decades, depending on the organism), and appropriate release of sperm for fertilization [132]. Beyond the mating plug, Sfps play an important role in all three stages of sperm storage [2, 133, 134].

In *Drosophila*, a few Sfps have been implicated in sperm storage [6-8]. These include the glycoprotein Acp36DE, an inferred proteolytic substrate (along with ovulin) of the seminal metalloprotease CG11864. Acp36DE is cleaved within the female reproductive tract after



mating, though the fates of individual cleavage products are not known, nor is the function of its cleavage.

Acp36DE is required for efficient storage of sperm [6, 7], possibly through its role in effecting post-mating conformational changes of the uterus [135]. Because females mated to males lacking Acp36DE store fewer sperm than mates of normal males, sperm from Acp36DE null males are less able to defend against displacement by sperm from a second male [136]. Acp36DE is also a component of the mating plug [137]. Whether any of these functions of Acp36DE are related to its processing by CG11864 has yet to be determined.

Several alleles of the *D. melanogaster* trypsin protease inhibitor Acp62F, another Sfp, are associated with variation in sperm competition success [79, 80]. Sperm competition refers to the ability of a male's sperm to resist displacement by a second male's ejaculate (defense) or to induce displacement of a previous male's ejaculate (offense). Complete deletion of the Acp62F gene results in males with improved defensive sperm competitive ability [80]. Unlike with Acp36DE, this does not appear to be related to the number of sperm stored. Instead, the improvement in "defense" may be a side effect of Acp62F's true, unknown function. For example, Acp62F may be involved in efficient release of sperm from storage; loss of this protein may make sperm harder to displace by rival sperm, but also less efficient at fertilizing eggs. Identifying the protease(s) inhibited by Acp62F may help shed light on its role in sperm use.

Though sperm storage and sperm competition are also important in mammals, these phenomena are much more difficult to study in mammalian systems than in insects. As a first step toward identifying Sfps with roles in these two processes, Claydon, *et al.* [96] studied the protein turnover rate of Sfps in the mouse (*Mus musculus domesticus*). A relatively high protein turnover rate is a prerequisite for a male's ability to control the abundance of specific Sfps in the

ejaculate in response to perceived sperm competition [138]. A high turnover rate would allow for a protein's abundance to be adapted on a short timescale. Six abundant seminal vesicle proteins were measured for incorporation of a dietary heavy isotope label over time, including SPIKL (a Kazal-type serine protease inhibitor), and all had higher than expected turnover rates of about 10% per day [96]. These data suggest that male mice may have the ability to change their ejaculate composition in response to social cues. This has been demonstrated in flies, as male *Drosophila melanogaster* are able to strategically allocate specific Sfps to females based on perceived levels of sperm competition [139]. The exact role of individual proteins, including protease inhibitors, in the context of sperm competition will require further study.

### ***Proteolysis regulators in sperm maturation and activation***

Many of the proteins important for sperm maturation and sperm activation (or capacitation) in mammals are not part of the seminal fluid *per se*, but rather are components of the sperm plasma membrane [140]. However, there are protease inhibitors in the seminal fluid that are important for mammalian sperm function [141]. Spink2 (for 'serine protease inhibitor Kazal-type 2') is expressed in the male germ cells of mouse and is required to protect the developing sperm against protease activity during spermatogenesis [142]. Thus, testis-expressed Spink2 may not be a secreted Sfp. However, it is also expressed in the mouse epididymis [142] and may play additional roles after spermatogenesis.

Several recent reports support a role of seminal fluid protease inhibitors in protection of the sperm from proteolysis. A seminal vesicle-secreted protease inhibitor, Spink3, binds to the plasma membrane of the apical hook of mouse sperm [12]. During transit through the female reproductive tract, Spink3 appears to have a protective protease inhibitor function against

female-expressed trypsin [12]. Spink3 may have also have a function unrelated to its ability to inhibit trypsin proteases. Incubation of GST-Spink3, which was unable to inhibit trypsin activity due to its fusion tag, prevented the acrosome reaction in capacitated sperm *in vitro* [143] by reducing the levels of sperm nitric oxide (NO) [143]. Very high levels of NO inhibit sperm function, though a moderate increase in NO production is required for full sperm maturation following capacitation [144, 145]. This non-inhibitory function of Spink3 may be required to keep sperm from completing the acrosome reaction until it reaches the ovum [143]. Another protease inhibitor, SPINKL (Spink-like), is secreted by the mouse seminal vesicle where it protects against premature capacitation [146]. Thus, protease inhibitors are important for regulating sperm capacitation, acrosome reaction, and female proteolysis.

Proteolysis regulators are required for sperm function in other organisms as well. A serine protease and two serine protease inhibitors are important for sperm activation in nematodes [11, 13, 147]. In the androdioecious species *Caenorhabditis elegans*, activation of male sperm occurs during insemination and involves a cellular rearrangement of the spermatid that confers motility on the mature spermatozoon [148]. Though males are not required for reproduction in *C. elegans*, since hermaphrodites can self-fertilize, males' sperm are preferentially used when mating has occurred [148]. A predicted serine protease inhibitor, SWM-1, is required to prevent premature sperm activation in males [147]. In the absence of SWM-1, sperm become activated in the male tract and cause infertility by preventing the transfer of sperm to hermaphrodites. The importance of a serine protease inhibitor suggests a role for a serine protease(s) in *C. elegans* sperm activation. This was recently confirmed [11]. A trypsin-like serine protease, TRY-5, is present in the ejaculate of *C. elegans* males and is required for sperm activation in the absence of hermaphrodite-specific activation factors [11]. Work in a related

nematode species, *Ascaris suum*, recently identified an additional serine protease inhibitor, As\_SRP-1, that is expressed in the spermatid itself and is required for the cytoskeletal changes at activation that lead to sperm motility [13]. Following sperm activation, As\_SRP-1 is secreted into the seminal fluid where it prevents activation of any sperm not already activated, presumably by blocking protease activity in the seminal fluid [13]. Taken together, these studies demonstrate a crucial role of proteolysis in nematode sperm activation, and that this activity must be highly regulated both in time and space.

Trypsin-like serine proteases are also implicated in sperm activation in insects. In *Manduca sexta*, sperm acquire motility in the spermatophore following ejaculation, in a process dependent on proteases secreted in the male reproductive tract [149]. *In vitro* experiments indicate that a trypsin protease(s) is involved in the activation of sperm motility. Similarly, a serine protease called initiatorin is required for sperm activation in another lepidopteran, the silkworm *Bombyx mori* [91, 150]. Initiatorin may function by activating a protein-arginine degradation cascade in the seminal fluid [150]. As evidence of this, purified recombinant initiatorin is proteolytically active and can activate sperm *in vitro* [91]. Serine proteases are also likely to function in sperm activation in the true bugs *Aquarius remigis* (water striders). Miyata *et al.* [151] found a crucial role for trypsin activity in initiating conserved signaling pathways for sperm motility in this species, though it is unclear whether the required trypsin protease(s) are present in the seminal fluid.

### ***Links between Sfp proteolysis regulators and infertility***

Recent studies support the use of proteolysis regulators in the seminal fluid as biomarkers for infertility. A study of 473 Han-Chinese men found significant correlations between alleles of

the epididymal serine protease inhibitor Eppin and some semen parameters, including sperm number, curvilinear velocity, and straight line velocity [152]. These parameters are used to assess the quality of semen in men seeking treatment for infertility. Eppin is part of a protein complex on sperm that includes lactotransferrin, clusterin, and SgI [reviewed in 153]. Binding of Eppin to SgI is critical to the role of SgI in inhibiting sperm motility prior to semen liquefaction [154]. In support of a role for Eppin in human fertility, treatment of male macaques with anti-Eppin antibodies resulted in complete, and reversible, loss of their sperm's motility [155]. Anti-Eppin antibodies from infertile monkeys also inhibit human sperm motility *in vitro* [156].

Studies of protein abundance in human semen have uncovered correlations between metalloproteases and semen parameters. Levels of the latent form of the matrix metalloprotease MMP-9 are elevated in semen samples with low sperm concentration [157]. MMP-9 and MMP-2 are also present in canine seminal fluid [158]. In dogs, ejaculate volume was negatively correlated with latent MMP-2 and latent MMP-9. Both the latent and activated forms of dog MMP-9 were negatively correlated with sperm concentrations, similar to the findings in humans [157]. The cause and effect relationship between these parameters remains to be determined.

Functional studies in model organisms will be crucial for determining how these proteins may affect fertility in humans. For example, the same study linking SERPINE2 in mice to mating plug malformation and infertility also compared the levels of the human homolog of SERPINE2 in semen samples between fertile and infertile men [130]. In contrast to the situation in mice, infertility in men with low semen fructose levels was correlated with higher than normal levels of SERPINE2. Though these results are in opposition to the role of this protein in mice, they nevertheless still indicate an important role of SERPINE2 in fertility.

### ***Protective effects of Sfp proteolysis regulators***

In mammals, Sfps are important in protecting sperm from attack by the female immune system and in promoting immune tolerance that is essential for pregnancy establishment [159]. Recent work has found direct links between seminal plasma and modulation of female immune function in mice [15, 16]. Seminal fluid serine proteases, specifically KLKs, have been implicated in activation of latent TGF-beta in the semen [160]. TGF-beta is secreted primarily from the prostate and additionally by other male accessory glands. It is thought to be the major contributor of seminal fluid immunosuppression, accounting for approximately half of the immune tolerance of the semen [161-163]. KLK14 can activate latent TGF-beta *in vitro* at physiologically relevant levels through cleavage of the N-terminal latency-associated pro-peptide (LAP) [160]. KLK14 may also cleave latent TGF-beta1-binding protein, contributing to TGF-beta activation. Emami and colleagues [160] demonstrated that KLKs 1, 2, and 5 may also be involved in TGF-beta activation by nicking the LAP and allowing for further activation by KLK14.

In addition to direct perturbations of female immune function, presence of antibacterial molecules in the seminal fluid of mammals [164] and insects [165, 166] argues that one function of Sfps is to protect against bacterial infection of the reproductive tract. This is important, since sperm may come into contact with microbes in the vagina introduced during copulation, which may cause harm within the uterus or fallopian tubes. Proteolysis of the seminal clot has recently been identified as a major contributor to the antibacterial activity of seminal plasma in humans [167]. Partially processed Semenogelin proteins had high levels of antibacterial activity, which was lost as the proteins became fully processed [167]. The timescale over which this happens mirrors that of the loss of antibacterial activity in whole semen [167].

In *Drosophila melanogaster*, ectopic expression of one metalloprotease Sfp (CG6168) or either of two serine protease inhibitor Sfps (Spn3 and CG42537—previously annotated as CG10284)—can protect females from systemic bacterial infection [65]. At least one of the protease inhibitors, Spn3, enters the female circulatory system after normal transfer from the male at mating [78]. This may be important in allowing the female to better fight off infections that might otherwise divert resources from egg production or decrease the female's lifespan. The added boost from Sfps may be especially beneficial, given that female production of antimicrobial peptides in response to infection is lowered after receipt of seminal fluid [168].

### ***Toxicity of proteolysis regulators in seminal fluid***

Identifying Sfps with toxic effects may help to get to the cause of certain pathologies in the male reproductive tract. Misregulation of proteolysis pathways in particular is likely to lead to damage of the male tract. Therefore, studying the regulation of Sfp proteolysis pathways and the consequences to their misregulation is a promising approach for further understanding of these pathologies.

In contrast to the protective effect of seminal liquefaction on bacterial infection, recent work has shown that Sg degradation by PSA during liquefaction may form amyloid fibers that increase the probability of HIV infection [169]. PSA also has a more direct deleterious effect—it is an allergen in women with seminal plasma hypersensitivity [170] and potentially also in men with post-orgasmic illness syndrome [171].

The toxic effects of some Sfps are well established in *Drosophila* [65, 77, 172]. Ectopic expression of several Sfps, including two protease inhibitors (Acp62F and Spn2), can cause lethality when expressed in preadult flies [65, 77]. One of these Sfps, the trypsin protease

inhibitor Acp62F, localizes to the uterus and sperm storage organs, while approximately 10% of this protein enters the female's circulation [77]. Though ectopic expression of Acp62 in adult virgin females did shorten their lifespan [77], females mated to males lacking this Sfp do not experience any change in lifespan [80]. This is likely because other factors (e.g., sex peptide [173]) overwhelmed the effect of Acp62F in the experiment. Identifying the targets of Acp62F will likely reveal why overexpression of this protease inhibitor causes toxicity in females.

## **FUTURE DIRECTIONS FOR STUDIES OF SEMINAL FLUID PROTEOLYSIS REGULATORS**

The proteomic studies outlined in the first section of this chapter [47, 48, 52, 57, 58, 83] have led to the identification of a large number of proteolysis regulators and their homologs in seminal fluid, but functions for the majority of those proteins are not yet known. Here, I attempt to outline some broad approaches that may lead to a better general understanding of the roles of proteolysis regulators in semen.

### ***Substrate profiling***

A first step to understanding what proteases are doing in the seminal fluid is to find what substrates they act upon. Identification of a protease as belonging to a specific class does not in itself suggest possible biological functions for that protease. Though broad substrate specificities are known for some protease families, these still need to be determined empirically for individual proteases. Several powerful methods, collectively called “degradomics” [174], now exist to profile substrate specificities (i.e., identify particular residues cleaved by the protease) and



identify native protein substrates for individual proteases.

Unbiased methods such as 1D and 2D gel-based approaches can identify large numbers of protease substrates, though not their exact cleavage sites. One such method is PROTOMAP [175]. PROTOMAP combines 1D PAGE with LC-MS/MS to identify protein size changes (indicative of proteolysis) across a wide dynamic range of potential substrate sizes. PROTOMAP is especially useful when large numbers of samples have to be compared, for example, when reconstructing a timecourse of proteolytic events.

Other methods directly screen for proteolytic events. N-terminal peptide identification methods [reviewed in 174] allow simultaneous identification of protease substrates and cleavage sites by taking advantage of the unique N-termini produced by proteolysis [176-178].

Recent studies using state-of-the-art mass spectrometry methods have been applied to blood plasma to identify peptides produced by proteolysis *in vivo* [179, 180]. These methods allow for *de novo* sequencing of peptides and thus identification of peptides with unexpected modifications (such as proteolysis by a non-standard enzyme) and yield high-quality peptides with low false-discovery rates [181]. These methods should be directly applicable to degradomics in seminal plasma.

Using a combination of the above degradomic methods, investigators in a number of systems could identify substrate specificities and substrate identities for proteases in the seminal fluid. They could also be used to identify proteolytic events immediately following ejaculation or over a longer timecourse in whole seminal plasma in order to identify new members of post-mating proteolytic pathways. Degradomics could also lead to the identification of biomarkers for prostate cancer, a disease already linked to high levels of the serine protease PSA [115].

### ***Model organism studies***

Questions regarding Sfp functions are not easily addressed using human samples, and this is no less true for proteases and protease inhibitors. Model organisms, such as the mouse and *Drosophila melanogaster*, can be used to address the roles of proteolysis regulators in the ejaculate. Single protein knockout studies are feasible in both organisms; these are most informative when the protein in question has a large fertility effect (such as SERPINE2, described earlier). However, many proteolysis regulators do not appear to have such large effects (e.g., Acp62F [80] and CG11864 [81] and there may be functional redundancy among proteases or inhibitors. Thus, multiple genes may need to be knocked out to study the function of a group of proteins, a task that is much more practical in *Drosophila* than in the mouse. In addition, mating context—such as female genotype or mating status, temperature, nutritional status, or possibly bacterial infection of either mating partner—may be important in the function of a particular Sfp. These questions require repeated, statistically powerful experiments that are difficult to perform using mammalian models (e.g., mouse or human) but are routine in *Drosophila*.

Additionally, structural analyses of specific proteases or inhibitors could also benefit from a model organism approach. Identifying crucial residues for function—including exosites (important regulatory sites outside of the catalytic site)—is easily carried out in *Drosophila*, while follow-up studies using mouse models and *in vitro* studies with human recombinant proteins will be important in establishing relevance in humans.

### ***Unexplored territory: Inactive protease homologs and non-inhibitory serpins***

An astounding result from *Drosophila* is the number of inactive homologs of serine

proteases and non-inhibitory serpins that are present in the seminal fluid (Table 1.3). The majority of the predicted Sfp serine proteases have mutations in essential active site residues. In addition, five out of the nine predicted serpins appear to be non-inhibitory based on the absence of a conserved inhibitory loop found in serpins [71]. Many non-catalytic protease homologs in the seminal fluid of *Drosophila* show signatures of positive selection [66], possibly reflecting newly acquired adaptive roles in fertility.

Inactive protease homologs may nevertheless play a role in proteolysis pathways. They may bind the substrates of other proteases without cleaving them, acting as inhibitors by competing with other proteases for substrates. Inactive protease homologs may also have regulatory roles. For example, the non-catalytic rhomboid protease homologs (iRhoms) are transmembrane proteins required for shuttling EGFR ligands to the ER-associated degradation pathway in *Drosophila* [182] and for shuttling the metalloprotease TACE out of the ER to the cell surface in mice [183].

In *Drosophila*, two secreted inactive serine protease homologs are part of post-mating reproductive pathways, though the molecular mechanisms of these proteins are still unknown. The inactive serine protease homolog CG9997, an Sfp, is part of the pathway that binds sex peptide to sperm and allows for the long-term maintenance of sex peptide's functions in the female [67, 68]. The second serine protease homolog, CG9897, is expressed in the spermathecae (female sperm-storage organs) [107]. Knock-down of this gene with RNAi can result in faster processing of the egg-laying hormone ovulin (a *Drosophila* Sfp mentioned earlier in connection with ovulation), though there is variation in this effect [107]. CG9897 is part of a three-member gene family located on chromosome 2R that includes another spermathecal protein (CG3834) and an Sfp (CG32833). These other gene family members may contribute to the variation in ovulin

processing rate observed with CG9897 knock-down.

Noninhibitory serpins have many roles in mammals outside of the seminal fluid [184]. Noninhibitory serpins act as molecular chaperones, hormone transporters, and tumor suppressors [184]. Whether any Sfp noninhibitory serpins, in *Drosophila* (see Table 1.3) or other organisms, share any of these functions should be investigated.

***Missing pieces of the puzzle: Female substrates/regulators/interactors of protease substrates***

Finally, it will be important to identify female protein interactors, substrates, and regulators of seminal fluid-mediated proteolysis. In *Drosophila*, many genes and proteins are differentially regulated in the female in response to mating [185-187]. At least 5 genes are specifically upregulated in the absence of the seminal fluid protease inhibitor Acp62F [188]. Thus, Acp62F may have an important function in pathways leading to regulation of those genes.

Female-expressed proteolysis regulators are good candidates for interactors of male-derived proteases and protease inhibitors, as they may act together in post-mating proteolysis networks. As mentioned earlier in this chapter, at least two protease pathways are known that begin in the male and are completed in the female: the seminase/CG11864 pathway in *Drosophila* [63, 81] and the seminal clot liquefaction pathway in humans [112]. These pathways, and others not yet identified, may require the action of female-derived proteases.

Another pathway mediated in part by seminase [discussed further in Chapter 2] leads to the binding of the Sfp sex peptide to sperm, which allows for its long-term storage and function in females [68]. One or more trypsin proteases are required to cleave sex peptide from sperm [189]. Given the long term action of this proteolytic activity, it is almost certainly due to female-expressed protease(s).

Many proteases have been identified as expressed in female reproductive tracts. In *Drosophila melanogaster*, secreted proteins in the spermathecae (one of the two types of sperm storage organs) are significantly enriched for serine proteases [190]. At least one of these proteases, Send2, is expressed in response to mating [191] and may function in sperm storage or regulation of proteolytic pathways. A mating-responsive female protease has also been identified in the mouse [14]. The metalloprotease MMP-9 is expressed at higher levels in the uteruses of mated females, and removal of male seminal vesicles prior to mating abolishes this increase [14]. As this increase in MMP-9 expression is coincident with neutrophil infiltration to the uterus, increased MMP-9 activity may be required for implantation success [14].

In the mosquito *Anopheles gambiae*, a genomic cluster of serine protease genes was recently reported as expressed in the female reproductive tract [192]. Evolutionary analysis of these genes revealed signatures of directional selection at specific codons. Rapid evolution of lineage-specific female reproductive-expressed proteases has also been observed in *Drosophila mojavensis* and its sister species, *D. arizonae* [70, 193]. Together, these results suggest that female protease genes could be evolving in response to interactions with male proteins. They may function to counter-act the effects of Sfps, mediate sperm-egg fusion through breakdown of plasma membrane components, protect sperm from proteolysis, regulate male-derived proteases, or activate protease cascades important in regulation of post-mating processes.

## CONCLUSIONS

The recent advances in proteomics of the seminal fluid have led to exciting discoveries that will drive a more detailed understanding of the roles of Sfps. However, identifying functions

for each Sfp is a challenge: many may have redundant functions, or their effects may only be visible in certain contexts. By taking a pathway-centered approach, we may better be able to define the broad roles of Sfps and identify patterns that span taxa, as well as species-specific functions. As the studies outlined here have shown, a focus on proteolysis regulators is a good first step to identifying Sfp-mediated pathways and their functions in fertility.

## **ORGANIZATION OF THIS THESIS**

The bulk of this thesis is focused on the function of proteases in the seminal fluid of *Drosophila melanogaster*. Chapter 2 describes the dual function of a serine protease Sfp, *seminase*, in the long-term response of females to mating and the proteolytic pathway that includes the astacin metalloprotease CG11864. Chapter 3 further dissects the CG11864 protease pathway by exploring the regulation of this protease by both the male and female. Chapter 4 focuses on the proteolysis of one of the inferred substrates of CG11864, the Sfp *ovulin*. Specifically, I ask whether there is genetic variation in the rate of *ovulin* processing, which could be used to identify additional factors required for proteolysis of *ovulin*. I change gears for Chapter 5 to discuss work in which I contributed to identification of new candidate Sfp genes, including proteolysis regulators, based on expression pattern correlations with known Sfp genes. Chapter 6 briefly summarizes the results of Chapters 2-5 and provides suggestions for future work related to these findings.

This thesis also contains three appendices. Appendix 1 outlines preliminary experiments done to determine the best system for *in vitro* expression of *ovulin* processing pathway components (*ovulin*, CG11864, and *seminase*) that will be of use for future experiments aimed at

developing an *in vitro* protease assay for ovulin. Appendix 2 contains a discussion of evolutionary dynamics of Sfps in *Drosophila* and how they reflect the “molecular social” interactions between males (through their Sfps) and females. This discussion is a subset of a published review article [25]. Finally, Appendix 3 describes a preliminary project, which explored the role of mating in regulating expression of genes involved in oogenesis.

## CHAPTER 2

# THE *DROSOPHILA MELANOGASTER* SEMINAL FLUID PROTEASE ‘SEMINASE’ REGULATES PROTEOLYTIC AND POST-MATING REPRODUCTIVE PROCESSES<sup>1</sup>

## INTRODUCTION

Proteolysis regulators are a component of the seminal fluid of many animal taxa, including insects [2, 22, 57, 83, 92, 194], fish [100, 101], birds [103, 195], and mammals [95, 112, 130, 140, 196, 197]. However, the mechanisms by which seminal proteases act, and most of the processes they affect, in mated females are poorly understood.

A mechanism by which proteases may effect physiological responses is through proteolytic cascades. Because most proteases are synthesized as inactive zymogens and require the removal of a short N-terminal sequence for activation [198], a protease cascade can be rapidly set in motion without new protein synthesis. For example, in mammals a seminal protease cascade activates the protease prostate specific antigen (PSA), in order to rapidly liquefy the seminal clot formed following ejaculation [reviewed in reference 112]. The action of PSA is regulated in part by the protease inhibitor PCI [reviewed in 114], which controls the timing and extent of liquefaction. Seminal clots are an important feature of the post-mating response in many animals [1, 130, 199].

---

<sup>1</sup> A version of this chapter was previously published as an article with the same title and is reprinted here with permission: LaFlamme, B. A., *et al.* (2012) *PLoS Genetics* 8(1): e1002435. K. Ravi Ram conducted the original RNAi screen and contributed to the results in Figure 2.7.



Given the prevalence of proteolysis regulators in seminal fluid, it seems likely that they are involved in other processes whose effects may extend past the first few minutes after mating. The study of seminal fluid protease functions would benefit greatly from a genetic approach. *D. melanogaster* provides an excellent system in which to study the roles of seminal fluid proteolytic proteins. Analysis of *Drosophila* seminal fluid proteins (SFPs) capitalizes on a wide range of available genetic tools, physiological and behavioral assays, and both a well-annotated genome and seminal fluid proteome. In addition, though individual SFPs, including proteases [66], are not generally well-conserved between distant taxa [29, 30], the biochemical classes into which SFPs fall are conserved between insects and mammals [1, 31], suggesting that mechanisms of action are likely to be conserved as well.

Approximately 18% of the proteins in the *Drosophila* ejaculate have been identified as predicted proteases or protease inhibitors [57, 200]. Mass spectrometry-based estimates indicate that the abundance of individual proteolysis regulators varies, with some among the most abundant proteins in the ejaculate (*e.g.* Acp62F) and others among the least abundant (*e.g.* CG10587) [57]. Most SFP predicted proteolysis regulators are either serine proteases or serine protease inhibitors with unknown functions [50, 55, 57], though a few other protease classes have also been identified [31, 56, 77, 81]. Proteolysis regulators have been identified as expressed in reproductive tract tissues of other insects including *Tribolium* [92] and mosquitoes [194] and directly in the ejaculates of mosquitoes [201]. In crickets, a predicted trypsin-like serine protease in the ejaculate is important for inducing egg-laying in mated females [22]. Though proteases are emerging as a common SFP class in animals, there have been no studies determining whether protease cascades (*i.e.* proteolytic pathways that require at least two

proteases in sequence) are a common regulatory mechanism for seminal fluid-mediated post-mating traits.

In *Drosophila*, transfer of SFPs from male to female during mating induces physiological changes in mated females [reviewed in reference 2]. Two of these changes are increased egg production and reduced receptivity to remating. These changes occur in two phases: short-term and long-term, both of which are necessary for optimal fertility. The short-term response (STR) occurs within 24 hours of mating and is solely dependent on the receipt of SFPs [202], including the prohormone ovulin [105], CG33943 [67], the sperm storage protein Acp36DE [7, 135], and the action of sex peptide (SP) that is not bound to sperm [203, 204].

Long-term persistence of post-mating changes (the long-term response, or LTR) requires SP and multiple other SFPs, and the presence of sperm in storage [204]. SP binds to sperm during mating. Cleavage by an unknown trypsin protease(s) is required to release the active portion of SP from sperm within the female [189]. SP is gradually cleaved from stored sperm during the approximately 2 weeks that they remain in storage. As long as SP is released into the female, she continues to lay eggs at a high rate and is more likely to reject courting males [189]. If SP cannot be released from sperm, the LTR does not occur [189].

Fertility defects arise if SP cannot bind to sperm in the mated female, or if it cannot be released from sperm. Sperm binding by SP requires the action of at least four other SFPs: the predicted serine protease CG9997, the Cysteine Rich Secretory Protein (CRISP) CG17575, and the gene duplicate pair lectins CG1652 and CG1656 [67, 68]. These four “LTR proteins”, together with SP, function in an interdependent network to bind SP to sperm as well as to localize each other to the seminal receptacle (SR), the major sperm storage organ of the female [68]. In this network, CG9997 is cleaved into a 36kDa protein in the male ejaculatory duct/bulb,

prior to transfer to the female and is required for the normal transfer of CG1652 and CG1656. CG17575 is required to localize CG1652 and CG1656 to sperm and the SR. This final step is then required for SP to bind sperm and accumulate in the SR. If any one of the four LTR proteins is absent, SP does not bind sperm. These SP-free sperm are still stored in normal numbers, but cannot be efficiently released from storage for fertilization past the first 24 hours after mating [68], because SP is also required for sperm release [69].

In addition to SP, two SFPs involved in post-mating traits are known to be cleaved following deposition into the female. The prohormone ovulin is initially cleaved at about 10 minutes after the start of mating (ASM) [106]. Ovulin is required for a maximal ovulation rate in the first 24 hours following mating [20, 105]. Processing occurs via three cleavage events from the N-terminus of ovulin that ultimately results in the production of one major cleavage product (approx. 25kDa) and three minor products (each 5kDa or smaller) [106]. Ectopic expression experiments have shown that both full-length ovulin as well as two C-terminal fragments, roughly corresponding to cleavage products of ovulin, are each able to independently induce ovulation in virgin females [108]. The glycoprotein Acp36DE is also cleaved within mated females [205], starting at approximately 20 minutes ASM, as detected by Western blot [6, 205]. Acp36DE is required for efficient sperm storage [6, 7]. This protein is responsible for the conformational changes of the uterus immediately following the start of copulation, which are thought to aid the movement of sperm into the storage organs [135, 206].

A previous study of 11 SFP proteases and protease inhibitors identified CG11864 as required for processing of both ovulin and Acp36DE [81]. Though all three proteins are produced in the male accessory glands, ovulin and Acp36DE are not cleaved until several minutes after their entry into the female reproductive tract. Therefore, three possibilities exist for

the regulation of ovulin and Acp36DE cleavage. CG11864 may be activated during mating, a repressor of CG11864 activity may be removed during mating, or a combination of both may occur.

CG11864 is predicted to be a member of the astacin family of metalloproteases, based on sequence similarity [81]. Astacin family proteases, like many other proteases, require removal of an N-terminal propeptide for activation [43]. The activity of CG11864 thus may be regulated in a similar manner. CG11864 is produced in the male accessory glands as a 33kDa protein and is cleaved to an approximately 30kDa form [81]. This cleavage begins in the male reproductive tract, in the ejaculatory duct and/or bulb, while CG11864 is in transit to the female during mating [81]. The size of the cleaved form of CG11864 is consistent with removal of a predicted propeptide from the N-terminus. We hypothesize that cleavage of CG11864 is required for its activation. If this is the case, there should be factors produced by the male, that regulate the activation of CG11864. However, the previous study involving 11 SFP proteolysis regulators did not suggest their requirement for the regulation of CG11864 [81]. A recent microarray analysis by Chintapalli et al. [62] and subsequent proteomic studies by Findlay et al. [57] identified additional serine proteases in the ejaculate. We, therefore, focused on these proteases to test for roles in the activation/regulation of seminal proteolysis.

Here, we used RNAi knockdown analysis to test five male-derived serine proteases for roles in ovulin cleavage and other reproductive events. We describe the first proteolytic cascade in fly seminal fluid that is regulated by a predicted trypsin-like serine protease, CG10586. We propose to rename this enzyme seminase (gene symbol: *sems*). Seminase is required for cleavage, and likely activation, of CG11864. Like CG11864, seminase is produced in the accessory glands and is cleaved in the male during copulation. We show that CG11864 is not

able to undergo self-cleavage in the absence of seminase. In addition to regulating CG11864 and thus its downstream SFP substrates, we show that seminase is a member of the LTR network, a CG11864-independent pathway that results in SP binding to sperm.

## **RESULTS**

### ***Seminase is required for normal CG11864-mediated processing of ovulin and Acp36DE***

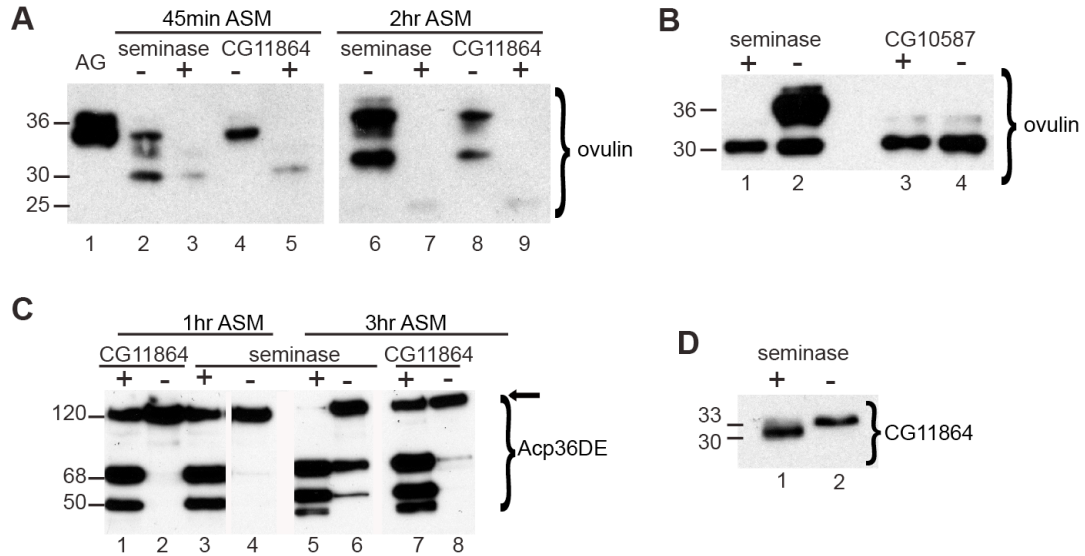
We tested five predicted protease SFPs for ovulin processing defects via Western blot to identify potential CG11864-interacting proteins. The tested SFPs were the predicted serine proteases 'seminase' (CG10586), CG10587, CG4815, CG12558, and CG32382 (sphinx2). Of these, only seminase, a predicted trypsin-type serine protease, was required for ovulin processing (Figure 2.1A).

Similar to the phenotype previously observed with CG11864 RNAi [81], some ovulin processing was observed in females mated to males knocked down for seminase, but ovulin was never processed fully in mates of seminase knockdown males, even at 2 hours ASM, the latest time at which ovulin can be reliably detected in female reproductive tracts when mated to controls (Figure 2.1A). Females mated to seminase knockdown males also failed to fully process Acp36DE at 1 hour ASM (Figure 2.1C), similar to CG11864 knockdown mates (Figure 2.1C). Even at 3 hours ASM, Acp36DE in females mated to seminase or CG11864 RNAi males had undergone only a small amount of processing relative to controls (Figure 2.1C).

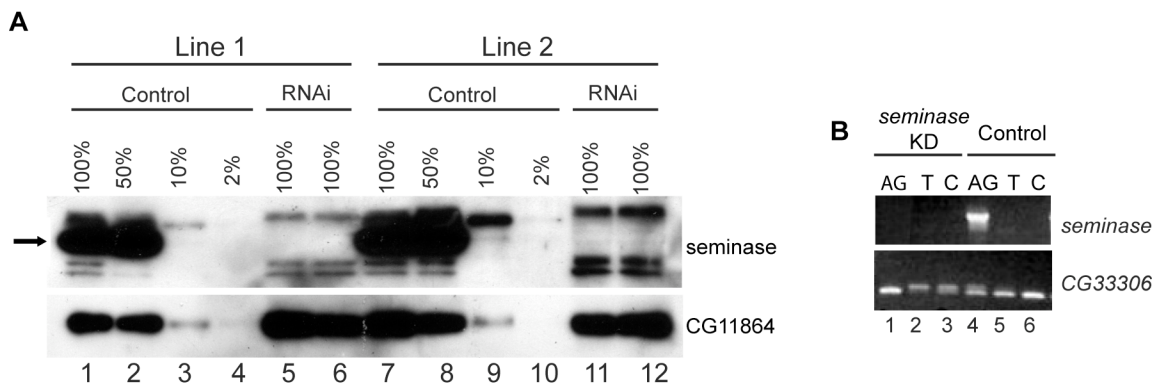
### ***Seminase is required for putative CG11864 propeptide cleavage***

Females mated to seminase RNAi knockdown males received CG11864 protein, but it was of the full-length molecular weight (33kDa); the cleaved form (30kDa) was never observed

(Figure 2.1D). However, females who mated to control males received both full-length and cleaved CG11864 (Figure 2.1D). Thus, *seminase* is required for the predicted pro-peptide cleavage of CG11864 during mating.



**Figure 2.1: Processing of SFPs is defective in the absence of *seminase*.** (A) Ovulin Western blot. Lane 1: full-length ovulin in male accessory glands (AG). Lanes 2-9: female reproductive tracts (RT) dissected after mating to control (+) or RNAi (-) males (genotypes given above lanes). Females were dissected at 45 minutes after the start of mating (ASM) (lanes 2-5) or 2 hours ASM (lanes 6-9). All lanes are from the same gel. (B) Ovulin Western blot. All: female RT at 30 minutes ASM. Lanes 1 and 2: females mated to *seminase* control (+) or RNAi (-) males. Lanes 3 and 4: females mated to CG10587 control (+) or RNAi (-) males. (C) Acp36DE Western blot. All lanes are from the same gel. Lanes 1-8: female RT, mated to males of the given genotype as in (A), dissected at 1 hour ASM (lanes 2-5) or 3 hours ASM (lanes 6-9). Un-processed (full-length) Acp36DE is ~122kDa. (D) CG11864 Western blot. RTs dissected at 45 minutes ASM. Numbers to the left of blots indicate approximate band size in kDa.



**Figure 2.2: Levels of seminase protein are reduced in RNAi lines.** (A) Western blot probed for seminase. Both independent insertion lines for the seminase RNAi construct are shown (Line 1 and line 2). Bands not affected by knockdown are assumed to be nonspecific cross-reactants. Arrow points to the seminase band. Each lane contains proteins from male accessory glands (AG). Lanes marked “100%” contain AG proteins from 10 males, “50%” from 5 males, “10%” from 1 male, and “2%” from the equivalent of 1/5 male. The bottom panel is the same Western blot probed for CG11864 as a loading control. Lanes 1-4: AG from line 1 control males. Lanes 7-10: AG from line 2 control males. Lanes 5-6: AG from line 1 RNAi males. Lanes 11-12: AG from line 2 RNAi males. (B) RT-PCR testing for presence of full-length seminase transcript (top panel) and a 200bp fragment of the CG33306 transcript (bottom panel). CG33306 is a potential off-target of seminase knockdown as predicted by the VDRC. Lanes 1-3 are from knockdown (KD) males. Lanes 4-6 are from sibling control males. Only results for Line 1 are shown, though similar results were obtained for Line 2. T: testes; C: carcass (whole male minus reproductive tract).

***Seminase is specific to the accessory glands and is initially cleaved in the male during transfer to the female***

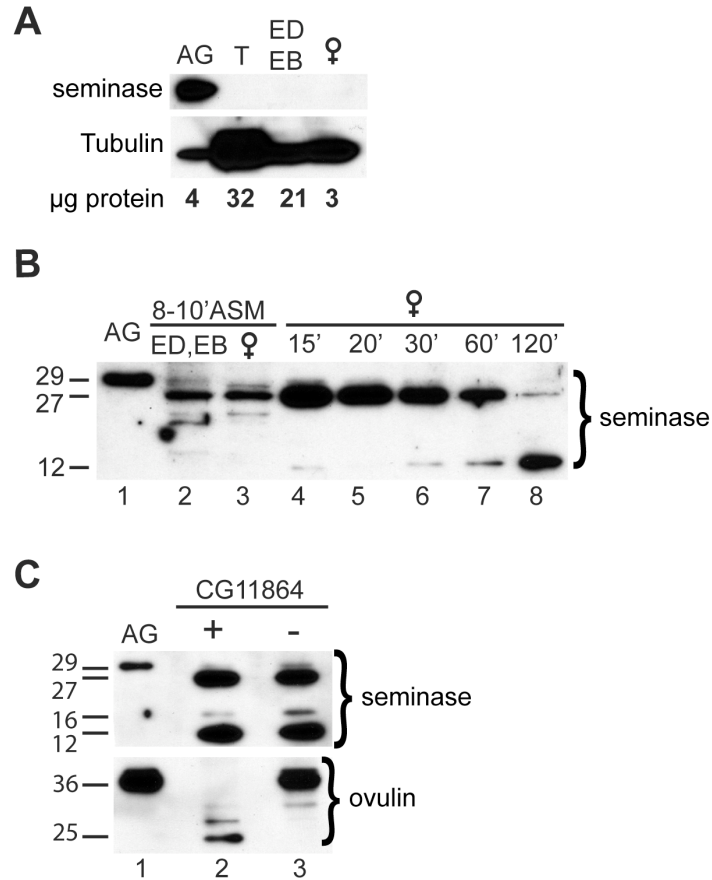
Since many serine proteases are synthesized as zymogens (containing an N-terminal sequence that must be removed for activation), we tested whether seminase was also processed during or after mating. Seminase is detected as an apparent 29kDa protein (predicted size: 28.2kDa, excluding a predicted N-term secretion signal sequence) in the accessory glands, with

no detectable expression in the testes or ejaculatory duct and bulb (Figure 2.3A). There was no evidence for seminase expression outside of the male accessory glands based on expression data in the FlyAtlas database [62] and our own RT-PCR (Figure 2.4B). We did not detect seminase protein in virgin females (Figure 2.3A).

During mating, an additional, lower molecular weight band (approximately 27kDa) of seminase appeared in the male ejaculatory duct and/or bulb (Figure 2.3B), consistent with removal of a 2.79kDa propeptide (size prediction based on an NCBI conserved domain search at [http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT\\_TYPE=live&SEQUENCE=NP\\_649270.1](http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE=live&SEQUENCE=NP_649270.1)). Within the mated female, seminase was further cleaved, producing a ~16kDa product (visible in top panel of Figure 2.1C) and increasing the amount of the ~12kDa product (Figure 2.3B). We first detected the ~12kDa product in the female at around 15 minutes ASM. However, due to the difficulty in detecting the ~16kDa form, we could not determine at what time ASM it is first produced.

Given that some proteases can be cleaved by their own proteolytic substrates [for example, see 207], we tested whether knockdown of CG11864 affected processing of seminase after mating. Females mated to CG11864 knockdown males showed normal processing of seminase at 30 minutes ASM (Figure 2.3C, top panel). As expected [81], CG11864 knockdown does prevent ovulin cleavage (Figure 2.3C, bottom panel), indicating a unidirectional proteolytic pathway.





**Figure 2.3: Seminase is produced in the accessory glands and is processed during and after mating.** (A) Western blot probed with seminase antibody. Seminase in the male accessory glands (AG) has an apparent molecular weight of ~29kDa. No seminase was detected in testes (T), the ejaculatory duct and bulb (ED/EB), or in virgin female reproductive tracts. AG and T: tissue from 10 virgin males. ED/EB: tissue from 20 virgin males. Female: tissue from 4 virgin females. Tubulin is shown as a loading control. Total protein loaded is shown in micrograms as measured by BCA assay. (B) Western blot probed with seminase antibody. Lane 1: full-length seminase in AG. Lane 2: ED/EB dissected from 20 males at 8-10 minutes ASM. Lane 3: Reproductive tracts (RT) from 20 females dissected 8-10 minutes ASM. Lanes 4-8: Female RTs dissected at the times ASM indicated above the lanes. (C) Western blots probed for seminase (top) and ovulin (bottom). Lane 1: full length protein in AG. Lanes 2 and 3: RT from females mated to CG11864 control (+) or knockdown (-) males at 1 hour ASM.

### ***Seminase is also required for CG11864-independent post-mating processes***

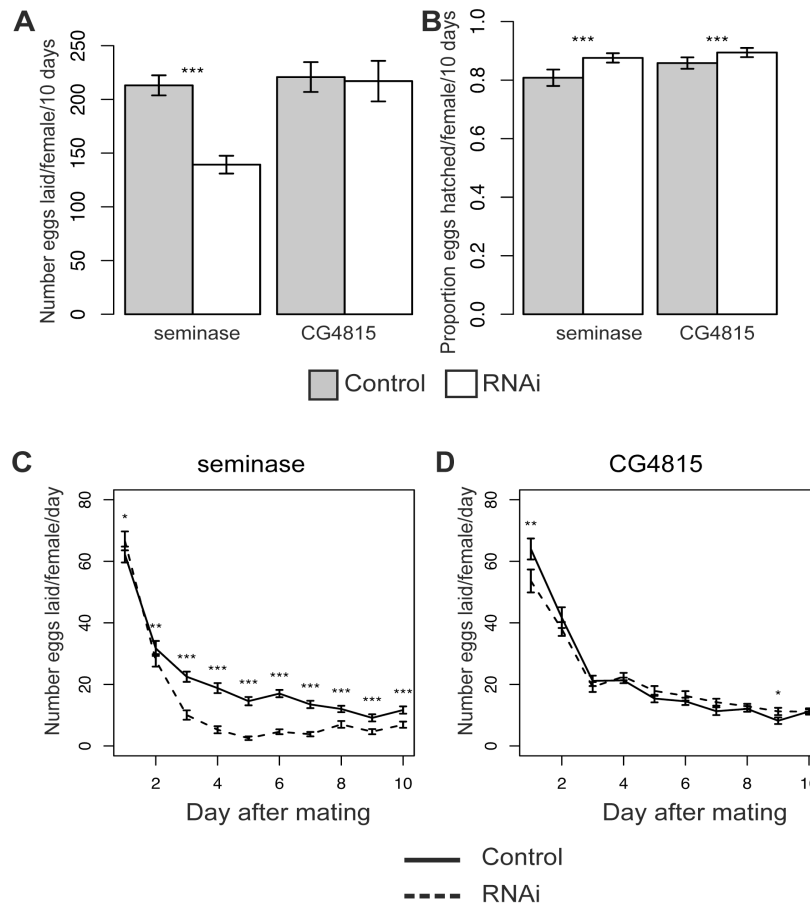
The total number of eggs laid over 10 days was significantly lower in females mated to seminase knockdown males relative to their controls, and this was seen in both independent insertion lines (Line 1 Poisson regression:  $z = -29.56$ ,  $p < 0.0001$ , Figure 2.4A; Line 2 Poisson regression:  $z = -31.59$ ,  $p < 0.0001$ , Figure 2.5A). There was no difference in total number of eggs laid between females mated to CG4815 knockdown and control males (CG4815 Poisson regression:  $z = -0.79$ ,  $p = 0.43$ , Figure 2.4A).

A repeated measures analysis of egg-laying over time revealed a significant effect of male genotype on egg-laying over time for both seminase lines and CG4815 (see Materials and Methods). To determine on which days male genotype affected egg-laying, data were analyzed separately for each individual day. The decrease in egg-laying, relative to control, in mates of seminase knockdown males was only apparent after the first day following mating and persisted until at least 9 days post-mating (Figure 2.4C and Figure 2.5C). Females mated to seminase knockdown males laid slightly, though significantly, more eggs than females mated to control males on day 1, but only with seminase Line 1 males (Figure 2.4C). These results indicated that seminase plays a role in long-term, though likely not short-term, egg-laying.

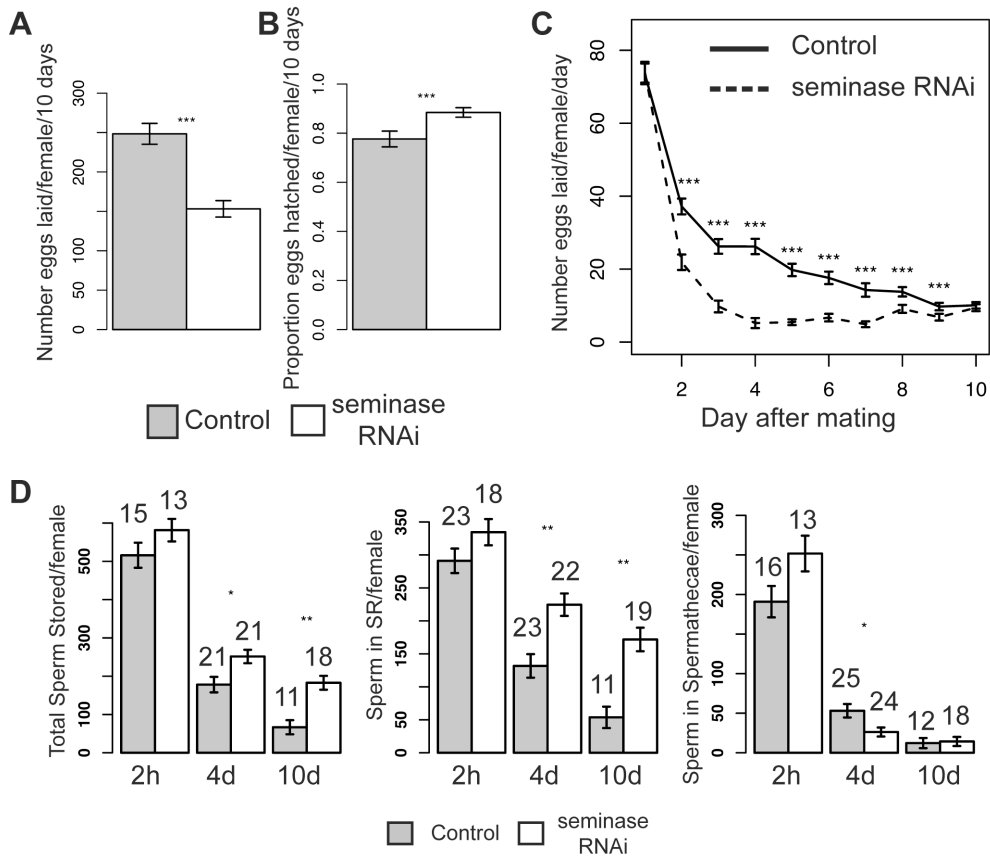
Females mated to CG4815 knockdown males laid significantly fewer eggs than females mated to control males on day 1 and slightly, though significantly, more eggs than controls on day 9 (Figure 2.4D). Thus, CG4815 may have a short-term effect on egg-laying, but not a long-term effect. These results also indicate that the long-term egg-laying effect of seminase is not an artifact of the VDRC strain background, which is shared by the CG4815 males.

Eggs laid by seminase or CG4815 knockdown mates hatched in significantly larger proportions than eggs laid by control mates (Binomial regressions: Line 1:  $z = 10.1$ ,  $p < 0.001$  and CG4815:

$z = 5.8$ ,  $p < 0.0001$ , Figure 2.4B; Line 2:  $z = 13.2$ ,  $p < 0.0001$ , Figure 2.5B), suggesting the egg-laying defect was not accompanied by a hatchability defect (hatchability is defined as the proportion of eggs that produced adult progeny), but rather that there was a slight deleterious effect of the balancer control background on hatch rates.



**Figure 2.4: Females mated to seminase knockdown males lay fewer eggs.** (A) The average number of eggs laid per female in a given treatment over 10 days. Seminase Line 1: Control N = 55, RNAi N = 59; CG4815: Control N = 20, RNAi N = 18. (B) Hatchability data for the corresponding experiments in (A). Hatchability is defined as the proportion of eggs that yielded adult progeny. CG4815 (a serine protease SFP) is included as a control for strain background. (A),(B) Asterisks indicate  $p < 0.0001$ . (C) The data from (A) plotted as average number of eggs laid by females in each group on individual days of the experiment. (D) CG4815 egg-laying data plotted as in (C). Asterisks indicate level of significance after Bonferroni correction (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.0001$ ). Error bars indicate standard error of the raw data pooled over experiments.



**Figure 2.5: Egg-laying, hatchability, and sperm storage in females mated to seminase Line 2 males.** (A) The average number of eggs laid per female in a given treatment over 10 days in seminase (Control N = 40, RNAi N = 49). (B) Hatchability data for the experiment in (A). (A),(B) Asterisks indicate  $p < 0.0001$ . (C) The data from (A) plotted as average number of eggs laid by females in each group on individual days of the experiment. Asterisks indicate level of significance after Bonferroni correction (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.0001$ ). (D) Sperm storage results for seminase Line 2, plotted as in Figure 4. Left panel: Total average number of sperm stored in both storage organs (2h:  $t = -1.45$ ,  $p = 0.15$ ; 4d:  $t = -2.5$ ,  $p < 0.05$ ; 10d:  $t = -4.4$ ,  $p < 0.001$ ). Middle panel: Average number of sperm stored in the seminal receptacle only (2h:  $t = -1.6$ ,  $p = 0.12$ ; 4d:  $t = -3.8$ ,  $p < 0.001$ ; 10d:  $t = -4.9$ ,  $p < 0.0001$ ). Right panel: Average number of sperm stored in the paired spermathecae only; numbers are the sum of sperm stored in each spermatheca (2h:  $t = -2.02$ ,  $p = 0.053$ ; 4d:  $t = 2.62$ ,  $p < 0.05$ ; 10d:  $t = -0.23$ ,  $p = 0.82$ ). Asterisks indicate level of significance. Error bars indicate standard error. Sample sizes are shown above each bar.

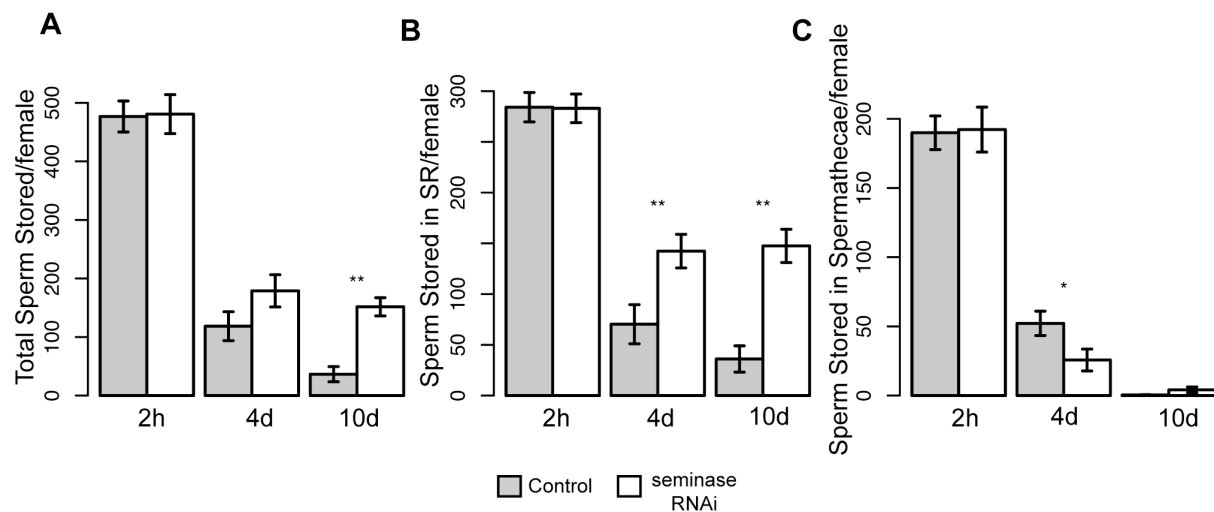
A reduction in fecundity after the second day post-mating suggests *seminase* is a new member of the LTR pathway. Increased recovery of post-mating receptivity to courtship beginning after the first 24 hours post-mating is also associated with LTR defects. Therefore, we tested whether lack of *seminase* in the ejaculate also caused increased receptivity in females, relative to controls, after mating. Table 2.1 shows data for female receptivity at 24 hours, 2 days, and 4 days ASM to *seminase* knockdown or control males. Similar to previously described phenotypes of LTR SFPs [67-69], females mated to *seminase* knockdown males were significantly more likely to remate at 2 days and 4 days ASM than were controls. Females mated to males from *seminase* knockdown line 1 showed a smaller magnitude of difference in remating rate relative to their controls than did line 2. This is most likely due to a background effect in line 1 that is apparent in the control males, as the remating rate is similar in mates to knockdown males from both lines. Since the same RNAi construct is expressed in both lines, we assume that the higher remating rate for females mated to line 1 control males is due to the insertion locus of the transgene.

**Table 2.1: Female receptivity to second mating.** Females were mated first to control or RNAi knockdown males from the line shown (either *seminase* line 1 or 2, or CG4815). CG4815 is included as a control for the VDRC strain background. The time given is the number of days after the first mating that the second mating was attempted. Females who remated to a male from a wildtype strain (Canton-S) during a one hour observation period are shown as a percentage with the total number of potential mating pairs assayed in brackets. P values in bold are significant.

Line	Time ASM	% Control Remated (N)	% RNAi Remated (N)	Chi Square	p value
<i>seminase</i> -1	24h	0% (29)	0% (25)	NA	NA
<i>seminase</i> -1	2d	5% (40)	57% (47)	24.44	<b>7.67E-07</b>
<i>seminase</i> -1	4d	42% (31)	94% (32)	17.19	<b>3.38E-05</b>
<i>seminase</i> -2	24h	0% (28)	3.3% (30)	0.001	0.972
<i>seminase</i> -2	2d	2% (47)	30% (46)	11.76	<b>6.06E-04</b>
<i>seminase</i> -2	4d	1.7% (57)	90% (52)	83.11	<b>7.74E-20</b>
CG4815	4d	0 (20)	0 (19)	NA	NA

### ***Seminase is required for sperm release from storage.***

Because LTR SFPs (in concert with SP) affect the release of sperm from storage [67, 69], we tested whether mates of seminase knockdown males also showed a defect in sperm release. Mates to seminase knockdown males stored normal numbers of sperm (Figure 2.6 “2h” bars), but significantly more sperm remained in storage at 10 days ASM in mates of seminase knockdown males than in control-mated females (Figure 2.6A).



**Figure 2.6: Females mated to seminase knockdown males retain more sperm 4 and 10 days after mating.** Sperm counts for seminase line 1 are shown for sperm stored at three timepoints after mating, as given on the x-axis. Line 2 yielded similar results. Asterisks indicate level of significance (\* $p < 0.05$ ; \*\* $p < 0.01$ ). (A) Total average number of sperm stored in both storage organs (2h:  $t = -0.09$ ,  $p = 0.93$ ; 4d:  $t = -1.6$ ,  $p = 0.13$ ; 10d:  $t = -3.2$ ,  $p < 0.01$ ). (B) Average number of sperm stored in the seminal receptacle only (2h:  $t = -0.06$ ,  $p = 0.96$ ; 4d:  $t = -2.8$ ,  $p < 0.01$ ; 10d:  $t = -3.2$ ,  $p < 0.01$ ). (C) Average number of sperm stored in the paired spermathecae only; numbers are the sum of sperm stored in each spermatheca (2h:  $t = -0.11$ ,  $p = 0.91$ ; 4d:  $t = 2.2$ ,  $p < 0.05$ ; 10d:  $t = -1.1$ ,  $p = 0.26$ ). Samples sizes are, for bars from left to right: (A) 12, 14, 16, 10, 7, 15; (B) 17, 21, 18, 19, 7, 15; (C) 13, 14, 19, 10, 7, 15. Abbreviations: 2h, 2 hours; 4d, four days; 10d, 10 days. Error bars indicate standard error.

This effect was due to a failure to release sperm from the seminal receptacle (SR) (Figure 2.6B), as sperm numbers in the spermathecae decreased at similar rates in females mated both to control and seminase knockdown males (Figure 2.6C). A slight, but significant, difference in

sperm release was seen in the spermathecae at 4 days ASM, but this effect was in the opposite direction from that seen in the SR and was no longer apparent by 10 days ASM (Figure 2.6C). Similar effects were seen with seminase knockdown Line 2 (Figure 2.5D).

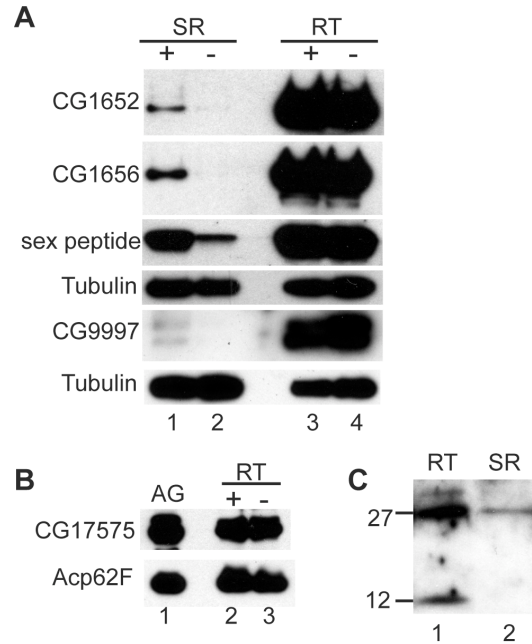
***SP and other LTR proteins fail to localize to the seminal receptacle in the absence of seminase<sup>2</sup>***

The results above are consistent with seminase being a member of the LTR network. To determine the placement of seminase in this network, we tested whether knockdown of seminase affected the post-mating localization of SP and the three LTR proteins that localize to the SR: CG9997, CG1652, and CG1656. At 2 hours ASM, seminase was required for accumulation of SP, CG1652, and CG1656 in the SR (Figure 2.7A). CG9997 was not detected in the SR at 2 hours ASM, so we tested females at 1 hour ASM. Seminase was also required for accumulation of CG9997 in the SR at this time point (Figure 2.7A). However, seminase was not required for proper processing of CG9997 or transfer of any LTR SFPs to the female during mating (Figure 5A, RT lanes), including CG17575 (Figure 2.7B).

A small amount of seminase also enters the SR (Figure 2.7C), suggesting that it could physically interact with other LTR proteins there. However, we were unable to determine whether other LTR proteins affected seminase localization to the SR due to the extremely low seminase signal within the SR. Multiple repetitions of the experiment failed to yield consistent results. As with previous efforts to detect LTR proteins in the spermathecae [68], we were not able to detect seminase in these organs (data not shown).

---

<sup>2</sup> With the exception of the CG9997 Western blot, the results presented in Figure 2.7A are the work of K. Ravi Ram.



**Figure 2.7: LTR proteins fail to accumulate in the seminal receptacle in the absence of seminease.** (A) Western blots probed for the proteins indicated to the left of blots. Lanes 1 and 2 contain 20 seminal receptacles (SR) dissected from females mated to control (+) or seminease knockdown (-) males at 2 hours ASM for SP, CG1656, and CG1652, and 1 hour ASM for CG9997. CG9997 was probed on a separate blot from CG1652, CG1656, and sex peptide. Lanes 4 and 5 contain 4 reproductive tracts (RT; not including ovaries or sperm storage organs, SSOs) from a subset of the females dissected in SR lanes. Tubulin is shown as a loading control for both gels. (B) Western blot probed for CG17575. Lane 1: male accessory glands (AG). Lanes 2 and 3 contain 4 RTs (excluding ovaries) from females mated to control (+) or seminease knockdown (-) males at 45 minutes ASM. Acp62F is shown as a loading control. (C) Western blot probed for seminease. Lane 1 contains female RT excluding ovaries and SSOs from wildtype mated females. Lane 2 contains 80 SR from wildtype mated females.



## DISCUSSION

### *An SFP, the predicted serine protease seminase, regulates processing of other SFPs through a proteolytic cascade*

To identify proteins that may interact with the predicted astacin-family protease CG11864 to process the SFPs ovulin and Acp36DE, we used RNAi to individually test five serine protease SFPs for ovulin processing defects. One of these proteins, the predicted trypsin-type serine protease ‘seminase’ (CG10586), is required for normal processing of ovulin as well as of the sperm storage protein Acp36DE. Because the phenotype of seminase knockdown was similar to that of CG11864 knockdown with respect to SFP processing, we hypothesized that both proteins might act in a single pathway. Additionally, because trypsin (serine) proteases are required for activation of some astacin-family proteases (of which CG11864 is one) [43, 208], we further hypothesized that seminase might act upstream of CG11864.

We therefore tested whether seminase regulates the cleavage, and thus activation, of CG11864. We found that seminase is required for the approximately 3kDa mobility shift of CG11864 that is seen in the male reproductive tract very soon after mating begins, suggesting that seminase may activate CG11864 by cleaving its propeptide. The apparent processing of pro-CG11864 by seminase and the subsequent processing of downstream substrates is suggestive of a proteolytic cascade. No such proteolytic pathway has, to our knowledge, previously been identified in insect seminal fluid. With the identification of this pathway in *Drosophila melanogaster*, we have found a model system for dissecting proteolytic pathways involving seminal fluid proteins that have consequences for fertility.

Proteolytic cascades typically have three steps [209]: 1) auto-activation of an initiator protease present in low amounts and triggered by an external stimulus; 2) activation of a more abundant propagator protease by the initiator protease; 3) activation of an executor protease by the propagator, which will cleave the downstream substrates. In addition, the propagator may also cleave, and thereby continue to activate, the initiator. Altogether, this causes a rapid propagation of the initial external signal.

While protein abundance is not necessarily related to potency, it is intriguing that seminease (the putative initiator) is relatively scarce in the ejaculate of *D. melanogaster*, which is consistent with the above model. Abundance estimates are based on the normalized spectral abundance factor (NSAF) obtained by mass spectrometry on mated females [57]. NSAF is an approximate measure of the relative abundance of a protein in a complex sample. Seminease ranks at 130 out of 138 (NSAF= $1.34 \times 10^{-4}$ ), with 1 being the most abundant and 138 the least [57]. CG11864 is similarly scarce (87/138; NSAF= $7.69 \times 10^{-4}$ ). This is in contrast to the much higher abundance of CG11864's substrates (ovulin: 20/138, NSAF= $1.2 \times 10^{-2}$ ; Acp36DE: 19/138, NSAF= $1.21 \times 10^{-2}$ ).

Also consistent with the protease cascade model, seminease is cleaved to a slightly smaller form during mating while still in the male reproductive tract. This may be an activational propeptide cleavage event, though this has not been directly tested. It is possible that seminease self-activates upon entering the ejaculatory duct, as is the case for many serine proteases [among tissue kallikrein pathways, for example see 112]. Our data suggest that seminease acts as the initiator in the cascade, and CG11864 acts either as the propagator, the executor, or both.

After transfer, seminease itself undergoes additional processing in the female (after the initial propeptide cleavage in the male), and these cleavage products may be important for the

function of seminase. On the other hand, they may simply be degradation products of seminase. However, both scenarios remain speculative.

We have shown that, in the absence of seminase, CG11864 is not cleaved to the predicted active form. However, the predicted propeptide cleavage site of CG11864, based on sequence threading to other astacin-family proteases [81], is not a trypsin site, as would be predicted if seminase were the only protease responsible for CG11864 activation. Interestingly, there are three trypsin cleavage sites present in the propeptide region of CG11864. It is possible that CG11864 is cleaved via a two-step mechanism (involving a trypsin and CG11864 itself), similar to what is seen for the propeptide cleavage of *Astacus astacus* (crayfish) astacin, the prototype of the astacin family [208]. Future studies using purified proteins *in vitro* will determine whether CG11864 is capable self-cleavage and whether seminase acts to directly cleave CG11864.

### ***Role of the seminase/CG11864 pathway***

Despite a severe delay in ovulin processing, knockdown of neither seminase nor CG11864 results in an egg-laying defect in the first 24 hours after mating [CG11864 data reported in 81]. This result is not surprising, however, given the rather small effect on egg-laying seen with a complete knockout of ovulin [20]. Additionally, ectopic expression of full-length ovulin is sufficient to induce ovulation in virgin females [108], suggesting that the additional effect of ovulin processing may be too small to detect with the current assay. It is also possible that, while seminase was knocked down to very low levels, there may still be sufficient seminase present for its role in early egg-laying.

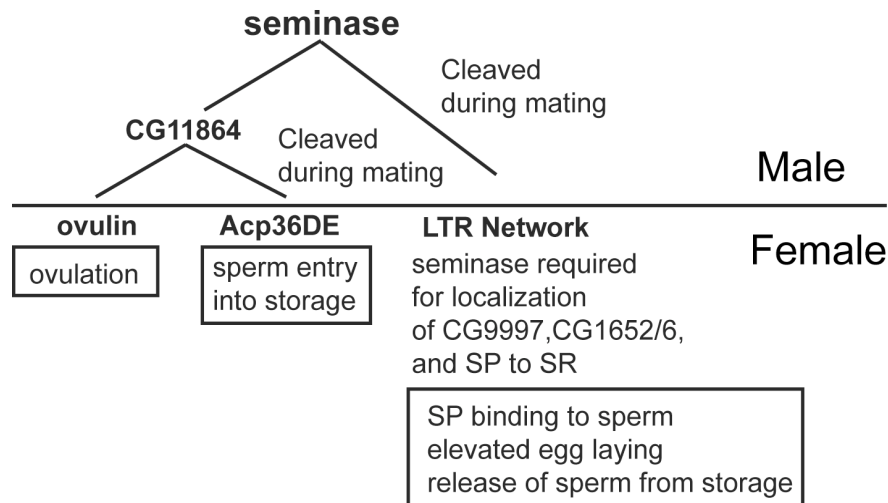
We also do not observe a defect in sperm entry into storage following seminase knockdown, as seen with knockout of Acp36DE [7]. However, Acp36DE processing may be

important for other functions of this protein. For example, Acp36DE is a component of the mating plug [137], but its function in the mating plug is still unknown. Further research is required to determine the consequences of loss of proteolytic processing of both ovulin and Acp36DE.

### ***Seminase regulates multiple independent post-mating processes***

In contrast to CG11864, which seems specific to the STR, seminase has a second important activity: it is in the LTR pathway, which regulates the binding of sex peptide (SP) to sperm [67-69] (See Figure 2.8 for overview). Similar to mates of SP null males, females mated to seminase knockdown males lay fewer eggs than controls over a 10 day period and also retain sperm in storage. While the interaction between sperm release and egg-laying is complex, over the long-term, egg-laying and sperm release are independent of each other [210]. Sperm do not directly influence the release of eggs, though the presence of SP bound to sperm is required for both sperm release [69] and normal post-mating levels of egg-laying [203, 204]. The failure of SP to accumulate in the SR indicates that seminase is likely required for SP to bind sperm.

The requirement for seminase in two independent post-mating pathways suggests that its activation at mating may act as a regulatory “switch” that coordinates post-mating events in *Drosophila*. Identification of other seminase substrates, if they exist, will allow us to determine the extent of seminase’s effects as a regulatory switch for post-mating events.



**Figure 2.8: Seminase is required for two post-mating pathways.**

Seminase is cleaved during mating within the male reproductive tract where it is required for CG11864 cleavage. CG11864 then regulates processing of two SFPs, ovulin and Acp36DE (left branch). Seminase is also necessary for the long term response (LTR) pathway (right branch). Proteins/pathways downstream of seminase are involved in the processes shown in boxes. The consequences of proteolytic processing of ovulin and Acp36DE remain unknown.

### *Evolutionary implications of seminase-regulated processes*

In the context of evolution, seminal fluid proteins represent a unique class of proteins in that they must, first and foremost, aid in successful fertilization, but are also tasked with representing the male's reproductive interests, sometimes in the face of opposing female interests [reviewed in 25]. This has the potential to set up a genetic conflict between the sexes and has been suggested to be one reason that seminal fluid proteins in particular tend to be rapidly evolving [29, 30]. However, an SFP's evolutionary rate is also likely to be constrained by the need for the protein to maintain its interaction with other proteins in the seminal fluid, as well as with proteins expressed by the female. For example, seminase and CG11864 are processed first in the male, but must interact with the female environment to further process both seminase and

the substrates of the proteolytic pathway. Seminase-regulated processes represent an opportunity to understand the evolution of seminal fluid protein networks that contain a mixture of conserved proteins (*e.g.* the lectins CG1652 and CG1656 and the CRISP CG17575 [57]) and proteins under positive selection (*e.g.* ovulin [211] and the serine protease CG9997 [212]). Seminase itself shows no evidence of positive selection, either at the protein level [212] or at individual sites (personal communication, Geoff Findlay). However, seminase does have two very closely related SFP paralogs, CG11037 and CG10587, which show evidence for recent positive selection in the *D. melanogaster* lineage [212]. These three genes are clustered together in the genome of *D. melanogaster* and the other *melanogaster* subgroup species. CG10587 does not play a role in ovulin or Acp36DE processing (CG11037 has yet to be tested), suggesting that these genes arose from tandem duplications and later diverged in function, with seminase remaining as the more conserved of the paralogs.

Our data on seminase show that this member of a conserved protein class in the seminal fluid plays a vital role in reproductive success. We believe that future study of the seminase-regulated pathways in *Drosophila* will lead to new mechanistic and evolutionary insights related to proteolytic cascades and protein networks in seminal fluid.

### ***Seminase-dependent processes: a model for pleiotropic seminal fluid proteins***

The regulation and mechanism of action of seminase constitutes a new *in vivo* model system for studying the regulation and physiological roles of pleiotropic seminal fluid proteins. Pleiotropic effects of kallikrein-related proteases involved in the liquefaction of human semen have recently been reported [160, 213]. Further understanding of the various effects of seminal fluid proteolysis in post-mating processes may have important implications for human health

(e.g. the role of PSA in cancer) and fertility. Our results indicate that genetic analysis in *Drosophila* will be an important complement to *in vitro* studies in mammalian systems for understanding the role of proteolytic processing in reproduction. Future studies of the regulatory mechanisms involved in the seminase/CG11864 proteolytic pathway may generate testable hypotheses for other seminal fluid protein networks, including those in mammals.

## MATERIALS AND METHODS

### *Flies*

Transgenic lines carrying RNAi constructs for *CG10586* ('seminase') and *CG4815* were purchased from the Vienna Drosophila Resource Center (VDRC, <http://stockcenter.vdrc.at>) GD RNAi library (P element library). VDRC lines used correspond to the following transformant ID numbers: 18795, 18796 (*CG10586*; same construct (ID 5539), different insertions), and 15410 (*CG4815*). *CG10586*-18795 is referred to here as "Line 1" and *CG10586*-18796 as "Line 2". *CG10586* VDRC lines are predicted to have an off-target, *CG33306*. However, we found no evidence for *CG33306* knockdown in either line (Figure 2.2B, only Line 1 shown). All RNAi knockdown and control sibling flies were produced by crossing *sympUAST-SFP* or VDRC virgin females to ubiquitous driver males (*Tubulin-Gal4 / TM3*). Non-Sb male progeny were knocked down for the SFP of interest and Sb flies were used as wildtype controls.

Matings were carried out by placing single 3-6 day old virgin females of the Canton-S strain with a single 3-6 day old virgin male in a glass vial containing a moistened square of filter paper. Matings were observed and the time at which mating began was recorded. Unusually short matings (<15 minutes) were discarded. Mated females were flash frozen in liquid nitrogen at the

appropriate time after the start of mating (ASM) for time points less than one hour ASM and stored at  $-80^{\circ}\text{C}$  until dissection. For time points greater than one hour ASM, females were frozen at  $-20^{\circ}\text{C}$  in glass vials and subsequently pooled and stored at  $-80^{\circ}\text{C}$  until dissection. All flies were reared in standard yeast-glucose media at room temperature ( $23\pm 1^{\circ}\text{C}$ ) on a 12:12 light/dark cycle.

### ***Seminase Antibody Production***

Generation of seminase fusion proteins and antibody purification was done following Ravi Ram et al. [78] and Cui et al. [214]. Briefly, we generated a 6xHis fusion protein containing amino acids 101-200 from seminase-PA using the pDEST17 vector of the Gateway system (Invitrogen). Antibodies were generated in rabbits (Cocalico) as described previously for eight other *Drosophila* reproductive proteins, including CG11864 [78], and Wisp [214] except that rabbits were immunized with the 6xHis-seminase fusion protein. Anti-seminase was affinity purified with a GST fusion protein of amino acids 101-200 of seminase, as described in the above references. Eluted antibodies were stored at  $-20^{\circ}\text{C}$  in glycerol (1:1), and used at a concentration of 1:250 for Western blot analysis.

### ***Sample Preparation and Western Blot Analysis***

Sample preparation and Western blot analyses in Figure 2.1A, 2.7A, and 2.7B were carried out as in Ravi Ram and Wolfner [68]. Samples in other Westerns in this study were prepared similarly, except that they were separated using 5-15% gradient SDS/PAGE. Female reproductive tract samples (RT) are lower reproductive tract extracts (ovaries removed) from 4-6 mated females, unless otherwise noted.



A BCA (bicinchoninic acid) assay (Pierce BCA Protein Assay Kit, Thermo Scientific) was performed to determine the protein loading in Figure 2.1A. Identical samples to those used for the Western were prepared and protein concentration measured relative to a BSA standard, following the manufacturers guidelines.

### ***Fecundity/Fertility Assays***

Number of eggs laid daily by mated females (fecundity) and number of progeny produced from those eggs (fertility) were quantified as in Ravi Ram and Wolfner [67]. Assays for the effect of seminase on fecundity and fertility were carried out three times for each independent insertion line, each time with 15-24 females measured for both knock down and control treatments. CG4815 knock down- and control-mated females were also measured for both fertility and fecundity, as a control for the VDRC background. Two assays were carried out with this line, each time with 7-12 females measured for each treatment. Hatchability was determined by dividing number of progeny by number of eggs (fertility/fecundity) [67]. Inspection of the data revealed no significant variation in egg-laying due to experimental block, so data were pooled across blocks. The effect of seminase or CG4815 knockdown on total 10-day egg-laying was tested with a Poisson regression model (using the R function 'glm()'. The statistical tests for hatchability were the same, except a Binomial regression was used.

A repeated measures analysis was performed to determine the effect of male genotype over time. This analysis was performed using a Poisson mixed-effects model with the R function 'lmer()' in the lme4 library. Two models were compared, a full model with day, genotype, and day-by-genotype interaction as fixed effects and female as the random effect, and a model with day as the only fixed effect. Comparison of the two models by ANOVA (R function 'anova()')

revealed the full model was the better fit, indicating a significant effect of male genotype. To determine the statistical significance of male treatment on individual days post-mating, a Bonferroni correction for multiple tests was applied to the Poisson regressions. All plots were generated using the means and standard errors of the raw data pooled from all experiments. Statistical significance of the effect of male genotype is denoted by asterisks on the plots.

### ***Receptivity Assays***

Females who had previously mated with either control males or seminase knockdown males were placed with a single wildtype male of the Canton-S strain for one hour either 24 hours, 2 days, or 4 days following the initial mating and the number of copulations beginning within one hour was recorded. Receptivity response to remating was tested for seminase as in Ravi Ram and Wolfner [67]. No fewer than 10 females were analyzed for control and experimental groups at 24 hours, 2 days, or 4 days after initial mating. Data were analyzed using a Chi-squared test (R function ‘chisq.test()’ with all parameters set to default).

### ***Sperm counts***

Sperm counts were performed as in Avila, *et al.* [69] at 2h ASM, 4 days, and 10 days ASM. Each slide was counted twice to assess counting precision (85%-100%). SR data used are the average of the two counts. Spermathecae data are the sum of the two averages (one for each spermatheca). Sample identity was coded to avoid bias. For 4 and 10 day samples, individual female daily egg counts were also taken. Numbers of stored sperm at 4 and 10 days ASM were significantly negatively correlated with the number of eggs laid (Figure 2.2). Females that laid very few eggs on day 1 (less than 2 standard deviations below the mean) were removed from the

dataset as they were likely unhealthy and may have had improper sperm storage. Data were analyzed using a two-tailed Student's t-test (R function 't.test()').

## **ACKNOWLEDGEMENTS**

We thank F.W. Avila for help with preparing samples for sperm counts and Norene A. Buehner for technical assistance with affinity purification of anti-seminase and with Western blots. We thank Jing Yang of the Cornell Statistical Consulting Unit for assistance with the statistical analysis.

CHAPTER 3

MECHANISTIC STUDIES OF THE CG11864-MEDIATED PROTEOLYSIS OF  
*DROSOPHILA MELANOGASTER* SEMINAL FLUID PROTEINS<sup>3</sup>

## INTRODUCTION

Seminal fluid proteins (Sfps) are important for male fertility in most species with internal fertilization, and in some with external fertilization [reviewed in 1, 2, 3]. Sfps function in sperm activation [4, 5, 10-13]; sperm storage in the female reproductive tract [6-9]; pregnancy establishment in mammals [4, 14-18]; ovulation in stimulated ovulators [20-23]; modulation of female immune function [15, 16, 19, 160]. They can even affect female post-mating behaviors [24, 25] and longevity [23, 172, 173] in some animals. In *Drosophila*, ablation of the major Sfp-producing cells of the male accessory glands (AG) results in complete loss of male fertility, unless Sfps are provided by a second male [215]. Over 200 Sfps are transferred (or inferred to be transferred) to females at mating in *D. melanogaster* [57-59, 200, 216]. However, functions have been assigned only to a relatively small proportion of Sfps in this, or any, species [for review, see 2].

The major protein classes of Sfps are common among distant taxa—from insects to humans—and include prohormones, lectins, cysteine-rich secretory proteins (CRISPs), proteases, and protease inhibitors [1]. Proteolysis regulators are a highly represented protein class in the

---

<sup>3</sup> The following contributions to this chapter were made by others: Dr. Frank Avila (Cornell University) assisted in generating the backcrossed CG11864 null stock and the CG11864 null *prd-Gal4* stock. Dr. F. Avila also contributed to results of fertility assays. Dr. Margarida Cardoso-Moreira (Cornell) analyzed sequence data from DGRP-517 to identify the location of the transposon insertion into the CG11864 locus.

seminal fluid of animals in which proteomic studies of Sfps have been carried out [this thesis, Chapter 1]. In humans, more than 100 proteins in the seminal plasma are proteolysis regulators (out of an approximately 950 proteins) [47, 48, 52]. In the model organism *Drosophila melanogaster*, proteins with predicted protease or protease inhibitor domains account for at least 14% of known or inferred Sfps [57-59, 200, 216], though individual proteins in this class tend to be of low abundance in the seminal fluid [57].

Despite the relatively large numbers of proteolytic regulators in the seminal fluid of many animals, this class of Sfps has not received much attention. Proteolysis regulators may be important in the regulation of other Sfps' functions, as proteases often act in cascades or proteolytic networks to rapidly amplify an external signal and activate downstream pathway components [209]. Understanding how proteases and protease cascades are regulated in the seminal fluid is of critical importance to studying the function of their substrates.

A single proteolysis cascade has been described in human seminal fluid. The chymotrypsin-like serine protease prostate-specific antigen (PSA) is activated by a complex proteolytic network involving several Kallikrein-like serine proteases [reviewed in 112][also see Chapter 1]. The function of PSA in this pathway is to degrade the seminal clot that forms quickly after ejaculation and that otherwise entraps and immobilizes a proportion of ejaculated sperm. There is also evidence for seminal fluid protease cascades in insects. In the silkworm *Bombyx mori*, the serine-type endopeptidase initiatorin appears to activate a protein-arginine cascade in the spermatophore [150]. This cascade may contribute to sperm activation in this species. In *Drosophila*, a proteolytic pathway is initiated in the male genital tract during ejaculation whereby the trypsin-like serine protease seminase is activated, allowing the rapid cleavage and activation of the astacin metalloprotease CG11864 [63]. CG11864, in turn, is required for

proteolytic cleavage of at least two downstream Sfps: ovulin and Acp36DE [63, 81]. Through a separate and still poorly understood pathway, seminease, along with several other Sfps including the predicted serine protease homolog CG9997 [67], is needed for the long-term function of the Sfp sex peptide (SP) [63], which is a major regulator of the female post-mating response [189, 203, 204].

The apparent activation of CG11864 by seminease provides a promising model for understanding the activation and function of seminal fluid protease cascades. Two proteases and multiple putative targets have been identified, providing ample opportunities for genetic manipulation and study. *Drosophila melanogaster* is an excellent model system for this approach, as targeted gene manipulation is routine in this species.

One predicted substrate of CG11864 is ovulin, which stimulates the release of eggs from the ovary during the first 24 hours after mating [20]. Ovulin is an Sfp prohormone that contains a short region of sequence similarity to the egg-laying hormone of *Aplysia californica* [217, 218]. Starting at approximately 10 minutes after the start of mating (ASM), ovulin is cleaved in three sequential steps from its N-terminal end to produce four cleavage products by 30 minutes to 1 hour ASM [106]. The function of ovulin cleavage remains unknown, though ectopic expression of the final two C-terminal cleavage products, as well as the full-length protein, can each independently stimulate ovulation in virgin females [108]. Whether the cleavage of ovulin *in vivo* constitutes degradation or activation remains to be determined. The results of Heifetz, et al. [108] suggest that the C-terminal cleavage products may be more effective than full-length ovulin at stimulating ovulation when ectopically expressed in virgin females, suggesting that the cleavage of ovulin may be activational.

Another predicted substrate of CG11864, Acp36DE, is a large glycoprotein Sfp that is

involved in sperm entry into storage [6, 7, 205], “offensive” aspects of sperm competition [136], and post-mating morphological changes to the female lower reproductive tract [135]. Acp36DE, like ovulin, is cleaved in the female reproductive tract [7, 205], and this cleavage requires CG11864 [81]. Acp36DE apparently undergoes only one cleavage, beginning at around 8 minutes ASM [81], to produce cleavage products of 68- and 50kDa from the 122kDa full-length protein [63, 81]. Further processing and/or degradation produces a product of slightly less than 50kDa 2-3 hours ASM [63]. As with ovulin, the role of Acp36DE cleavage in its activity is not yet known.

The role of CG11864 in the processing of downstream Sfps was dissected using RNAi knockdown of CG11864. Knockdown of CG11864 in males resulted in incomplete and drastically delayed cleavage of both ovulin and Acp36DE in mated females [81]. Because cleavage of these Sfps was delayed, but not completely lost, it was therefore not clear whether there was a second, partially redundant cleavage pathway for ovulin and Acp36DE processing, or whether expression of CG11864 was simply insufficiently knocked down.

Here, we report a null allele of CG11864 and demonstrate that in mates of null males there is a complete absence of processing of either substrate protein. The null allele was identified in one of the *Drosophila* Genetic Reference Panel (DGRP) inbred lines [219]. Though both male- and female-derived factors are required for Sfp proteolysis [106], the exact nature of the contribution from each sex is not clear. To further clarify the contributions from the male and female to Sfp proteolysis, we used our CG11864 null allele to test if CG11864 can be activated in females. We found that it can, without synthesis in or transit through the male reproductive tract. We then investigated the requirement of female factors for CG11864 activation using ejaculates collected from males. We found that the female reproductive tract is not strictly

necessary for CG11864 activity, though it is required for efficient processing of ovulin and Acp36DE.

Finally, we determined the functional significance of two conserved regions of the CG11864 protein—the predicted pro-peptide cleavage site and a structurally important catalytic residue—in its activation and activity. The results of this study further our understanding of a proteolytic cascade in *Drosophila* seminal fluid and will aid future research into the functional consequences of this cascade.

## RESULTS AND CONCLUSIONS

### *Identification of a null allele for CG11864*

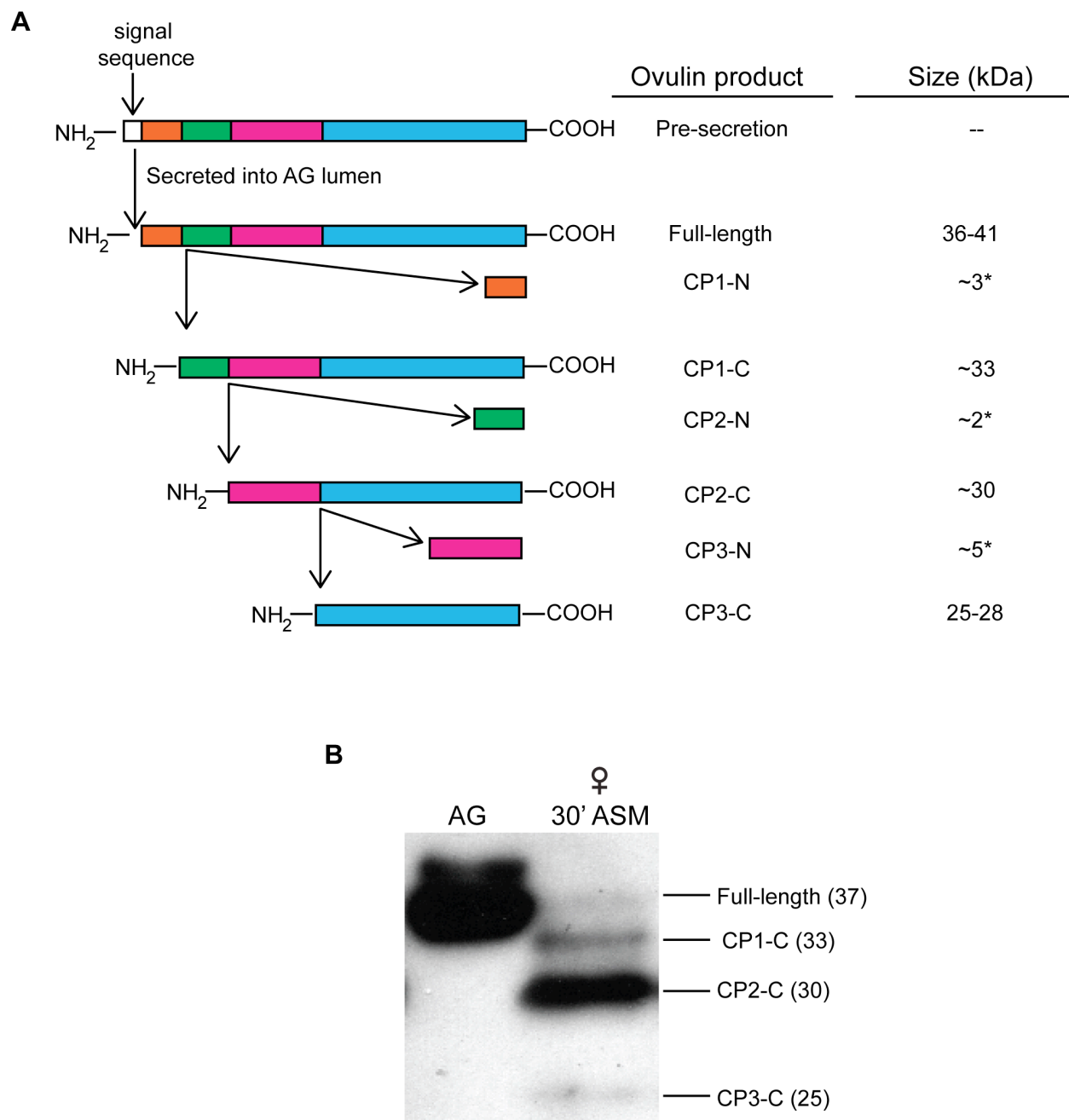
We screened through 35 of the DGRP inbred lines [219] by testing within-line matings for ovulin processing defects (Table 3.1). We found that ovulin was not cleaved in matings between males and females of one line, DGRP-517.

We detected cleaved ovulin in the reproductive tracts of wild-type (Canton-S strain, or CS) females mated to CS males at 1 hour after the start of mating (ASM) (Figure 3.2A, lane 1; a schematic of normal ovulin processing is shown in Figure 3.1.). In contrast, DGRP-517 females mated to males from the 517 line have only full-length ovulin at this same time ASM (Figure 3.2A, lane 4). To determine whether the failure to process ovulin was due to an abnormal contribution from the male, from the female, or from both, we mated CS females to DGRP-517 males (Figure 3.2A, lane 2) and vice-versa (Figure 3.2A, lane 3). We found that the DGRP-517 males alone are responsible for the ovulin processing defect.

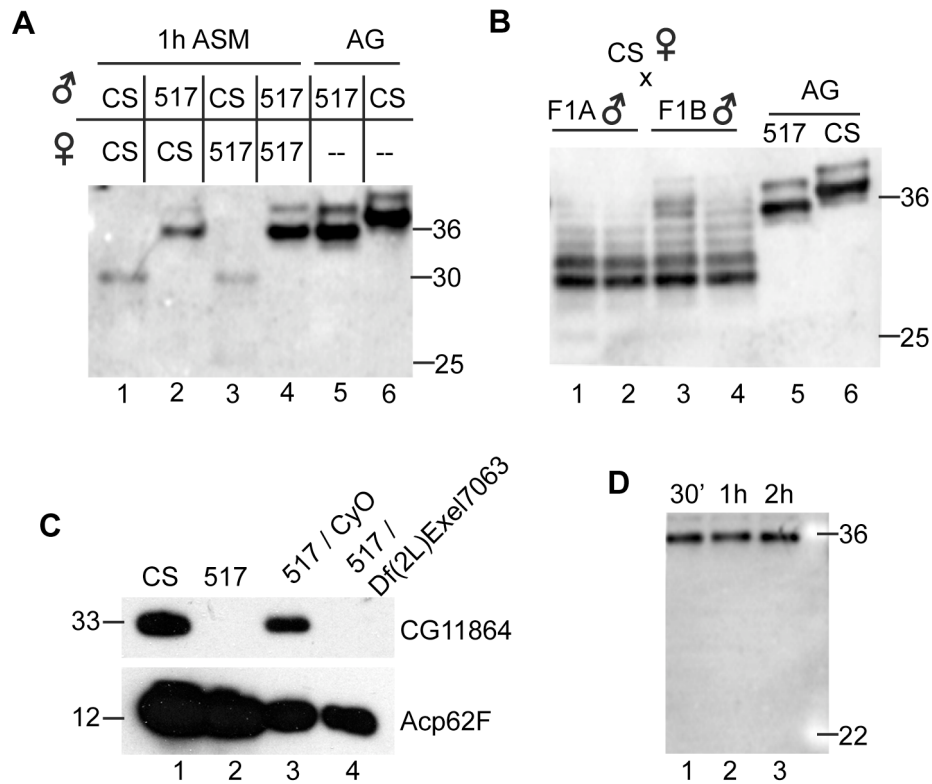


**Table 3.1: DGRP lines screened for ovulin processing defects.** Males and females from the same inbred line were mated and frozen at 30' ASM. Western blots of proteins from mated female reproductive tracts were probed with anti-ovulin antibodies and presence or absence of ovulin cleavage products was noted. Only DGRP-517 showed no ovulin cleavage products at 30' ASM.

DGRP line tested	Ovulin processed at 30' ASM?	DGRP line tested	Ovulin processed at 30' ASM?
28	yes	324	yes
45	yes	357	yes
49	yes	358	yes
57	yes	365	yes
59	yes	379	yes
91	yes	437	yes
101	yes	486	yes
161	yes	514	yes
176	yes	517	<b>NO</b>
208	yes	555	yes
272	yes	765	yes
301	yes	712	yes
303	yes	730	yes
304	yes	774	yes
315	yes	786	yes
319	yes	799	yes



**Figure 3.1: Schematic of ovulin cleavage pattern.** **A)** Ovulin is synthesized as a 264 amino acid precursor containing an N-terminal secretion signal. It is secreted into the lumen of the male AG after removal of the secretion signal peptide to the “full-length” protein. Full-length ovulin is transferred to females at mating and sequentially cleaved to cleavage products (CP) with approximate sizes shown to the right. Size ranges represent variation in ovulin glycoforms. Larger bands often appear as doublets. Asterisks denote cleavage products smaller than the cutoff used for gels in this study. **B)** Western blot probed with anti-ovulin showing full-length ovulin in male AG and mated female at 30’ ASM. All cleavage products are also visible in mated female lane.



**Figure 3.2: DGRP-517 males' seminal fluid lacks ovulin-cleaving activity linked to CG11864.** (A) Western blot probed with anti-ovulin antibody. Lanes 1-4: 4 female lower reproductive tracts (RT; uterus, sperm storage organs, and common oviduct) at 1 hour after the start of mating (ASM). Fly lines used for males and females are shown above blot. CS: Canton-S (wildtype strain); 517: DGRP-517. Lanes 5-6: 1 pair male accessory glands (AG) in each lane. (B) Western blot probed with anti-ovulin antibody. Lanes 1-2: 4 CS female RT (each lane) after mating to F1 males from 517 female x CS male cross. Lanes 3-4: Same as lanes 1-2, except using F1 males from the reciprocal cross. All mated female RTs are at 30 minutes ASM. Lanes 5-6: 1 pair AG each lane. (C) Western blot probed with anti-CG11864 (top) and Acp62F (bottom) antibody. Acp62F is used as a loading control. Each lane contains 2 pairs AG from males with genotypes shown above blot. (D) Western blot probed with anti-ovulin antibody. Each lane contains RT from 4 mated females after mating to DGRP-517 males. Lane 1: 30 minutes ASM; lane 2: 1 hour ASM; lane 3: 2 hours ASM. Numbers next to blots indicate approximate molecular weight in k-Da.

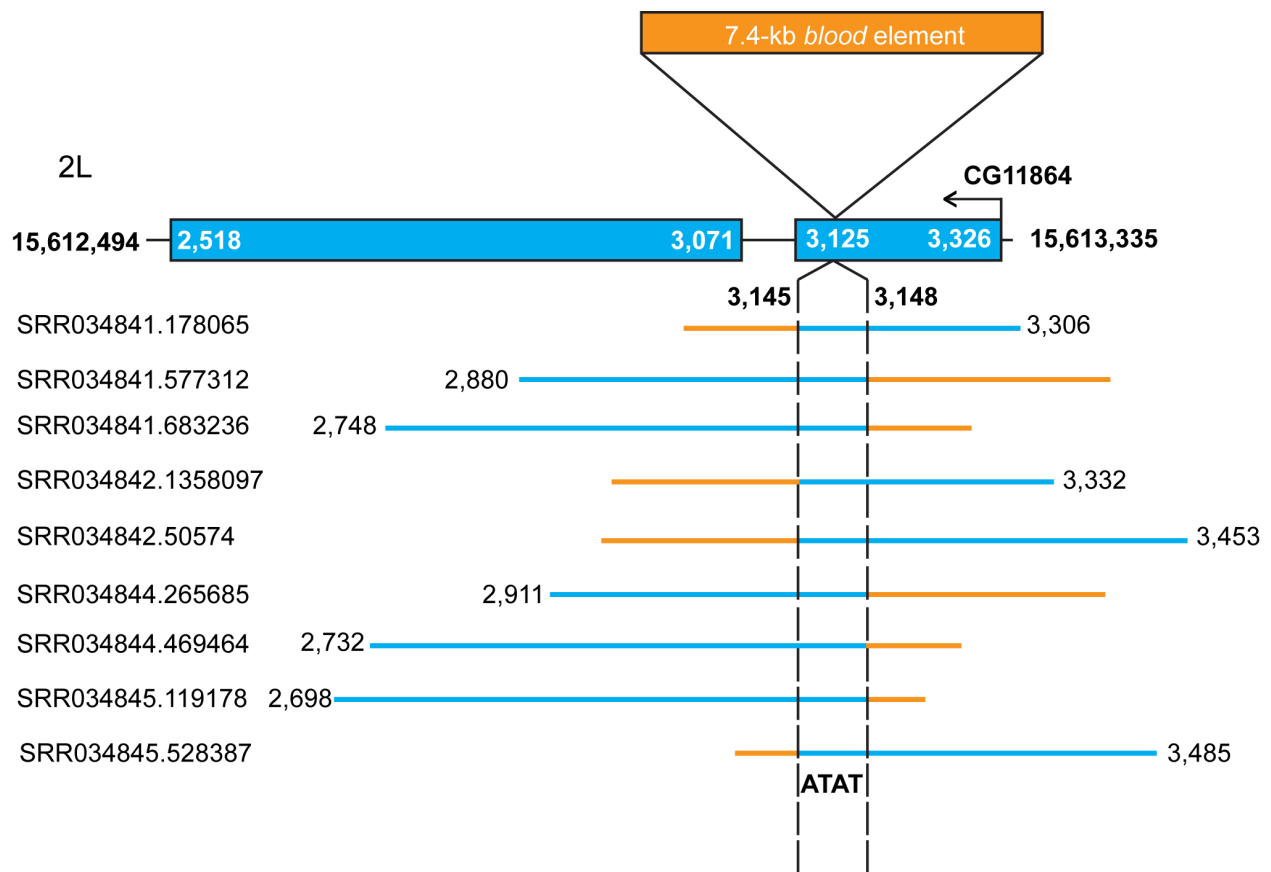
Ovulin from DGRP-517 males had a lower apparent molecular weight than did ovulin from CS males (Figure 3.2A). We examined the publicly-available sequence from the 517 line and determined that there is a small (7 residue) C-terminal truncation of ovulin due to the introduction of a premature stop codon. We confirmed this mutation in our own lab stock of DGRP-517. To determine whether the ovulin truncation is responsible for its failure to be cleaved in crosses involving DGRP-517 males, we crossed DGRP-517 females to CS males to obtain “F1A” males, and crossed CS females to DGRP-517 males to obtain “F1B” males. These males were then mated to virgin CS females to determine whether ovulin could be cleaved. If the cleavage defect is linked to the DGRP-517 ovulin allele, we would expect to see normal cleavage products from the CS allele, and full-length ovulin corresponding to the truncated allele. However, we found both isoforms of ovulin were processed by 30 minutes ASM (Figure 3.2B). Each cleavage event resulted in cleavage products of two sizes, one from each of the two ovulin isoforms. Thus, the defect in ovulin cleavage in DGRP-517 is not due to its ovulin sequence, but rather to a recessive, trans-acting factor.

The astacin-family metalloprotease CG11864, a seminal fluid protein (Sfp), is required for normal processing of ovulin [81]. We therefore checked whether DGRP-517 males produce normal CG11864. We were able to detect CG11864 on Western blots of male accessory glands (AG) from CS males (Figure 3.2C, lane 1). However, we could not detect CG11864 protein in AG from DGRP-517 males (Figure 3.2C, lane 2). To determine whether the lack of CG11864 protein is due to a defect at the CG11864 locus or to an unlinked locus, we tested hemizygotes for the CG11864 locus. Males with DGRP-517 chromosome 2L over a balancer chromosome with a WT allele of CG11864 had normal CG11864 protein levels (Figure 3.2C, lane 3), whereas males hemizygous for the DGRP-517 allele had no detectable CG11864 (Figure 3.2C, lane 4).

All males had normal levels of an unrelated Sfp, Acp62F, which was used as a loading control. We therefore concluded that DGRP-517 males have a molecular lesion linked to the CG11864 locus that results in loss of CG11864 protein.

No ovulin cleavage products were ever observed in females mated to DGRP-517 males, even at 2 hours ASM (Figure 3.2D). Similar results were obtained with Acp36DE (data not shown). This result is in contrast to previous work in which RNAi knockdown of CG11864 delayed, but did not fully preclude, proteolytic processing of these substrate proteins [81]. Results obtained with CG11864 RNAi males were not able to clarify whether the partial processing of ovulin and Acp36DE in females mated to RNAi males was indicative of another protease's involvement, or rather of incomplete knockdown of the CG11864 gene product. Ovulin is never fully processed in females mated to CG11864 RNAi males [81], but its first and second cleavage events do occur. Our data suggest that the ovulin processing seen with the RNAi males was due to incomplete knockdown of CG11864, and that CG11864 is essential for any ovulin processing to occur.

We were unable to amplify the CG11864 coding region using PCR from DGRP-517 flies (data not shown) despite the fact that no major mutations were apparent in this region of the DGRP-517 genome assembly. We then analyzed the raw sequence reads to determine the nature of the molecular lesion (with significant help from M. Moreira). We found a number of reads that map to the CG11864 locus at one end and to a transposable element (*blood* family) at the other end (Figure 3.3). From these reads, we can conclude that an approximately 7.4-kb *blood* element has been inserted into the first exon of CG11864 in this line, resulting in loss of the CG11864 gene product. The presence of the *blood* element was confirmed by PCR (not shown).



**Figure 3.3: Genomic location of *blood* element insertion in DGRP-517.** A 7.4-kb *blood* transposable element is inserted in the coding region of CG11864 at genomic location 15,613,145-148. The CG11864 gene region is located on chromosome 2L at 15,612,494...15,613,335 and transcribed from the minus strand. Here, the coding sequence of CG11864 is represented by blue rectangles with exon boundary locations in white numerals. 5' and 3' UTRs and the single intron are represented as solid black lines. 454 reads spanning the insertion site are shown: blue lines represent sequence mapping to the reference genome and are aligned with their approximate locations in the CG11864 genomic region. Orange lines represent the portion of each read that maps to the insertion and do not align with the CG11864 genomic region in the reference genome. Each read is marked with the physical location of the end that maps to the genome. The region of CG11864 between the dashed lines shows the inferred insertion site to within 4bp. The sequence at these bases is shown (ATAT); the ambiguity in the exact insertion site is likely due to the repetitive nature of the sequence surrounding the insertion. Figure is not drawn to scale.

***Males with the DGRP-517 allele of CG11864 in a wild-type background retain the ovulin processing defect***

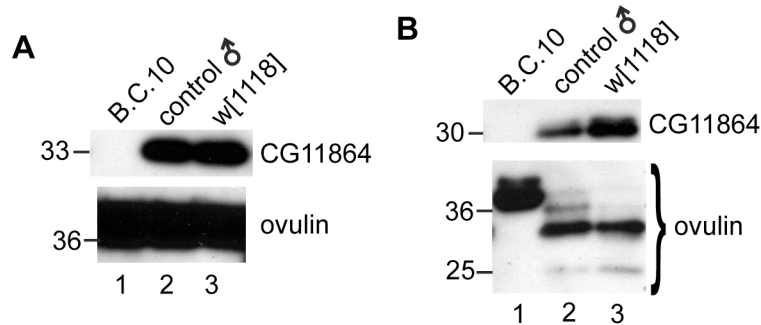
We isolated the DGRP-517 allele of CG11864 from the rest of the DGRP-517 genome using the crossing scheme outlined in the Supplementary Methods. This was necessary to remove any potential effects on ovulin processing from other mutations that might be present in the 517 line. The CG11864 null allele was first linked to a P-element containing a visible marker, so that flies carrying the null allele could be crossed into a standard white-eyed laboratory strain ( $w^{1118}$ ) for several generations. This also provided a good genetic control for phenotypic assays: we crossed the P-element-only line to  $w^{1118}$  for the same number of generations (“control males”).

We tested males after ten generations of backcrossing (Figure 3.4A) for the presence of CG11864 protein. Backcrossed males (“B.C.10” males) failed to process both ovulin (Figure 3.4B, lane 1) and Acp36DE (data not shown), another known substrate of CG11864 [81]. Control males produced similar levels of CG11864 and ovulin to  $w^{1118}$  males (Figure 3.4A, lanes 2 and 3). Control males also processed ovulin at a similar rate to  $w^{1118}$  males (Figure 3.4B, lanes 2 and 3).

***Ectopic expression of CG11864 from CG11864 transgenes in males and females***

We created a CG11864 transgene under the control of the GAL4 upstream activating sequence (UAS) in CG11864 null males. The full coding region (excluding the single intron) of CG11864 (Figure 3.5A) was inserted into chromosome 2L using the PhiC31 integrase system (see Methods). CG11864 null males carrying the transgene and expressing the *paired*<sup>GAL4</sup> driver [220] expressed CG11864 in their accessory glands, though at slightly lower levels than in CS

males (Figure 3.5B, lane 2). The balancer sibling controls of these males (CG11864 null males carrying the transgene but not expressing the driver) did not express CG11864 (Figure 3.5B, lane 3).



**Figure 3.4: Backcrossed CG11864 null males retain processing defects.** Western blots probed with anti-CG11864 (top) and ovulin (bottom) antibodies. **(A)** Lane 1: 10 pairs AG from CG11864 null males backcrossed 10 generations to  $w^{1118}$ , followed by sibling crosses to recover homozygous CG11864 null chromosome 2. Lane 2: 10 pairs AG each from control males backcrossed 10 generations to  $w^{1118}$ , followed by sibling cross. Lane 3: 10 pairs AG from  $w^{1118}$  males. **(B)** Lane 1: 6 mated female RTs 30min ASM to indicated males.

We also created two transgenes to ectopically express mutant forms of CG11864 in males. We mutated the two residues flanking the predicted CG11864 cleavage site (G45 and I46) to alanine (Figure 3.5A). This cleavage site is predicted to be necessary for CG11864 activation. We refer to this transgene as “cleavage mutant”. We also mutated a conserved residue within the catalytic domain, Q155, to glutamate. This residue is nearly universally conserved as glutamine in all known astacin-family proteins other than CG11864 [221]. Mutation of this residue in the canonical astacin protein (Glu103) to glutamine results in reduced stability of the protein [208]. All transgenes were expressed in male accessory glands of CG11864 null males under the



control of the *paired-GAL4* driver. In addition to the expected CG11864 band, we also observed a slightly higher molecular weight band on CG11864 Western blots in many males (Figure 3.5B). However, this does not appear to affect the function of CG11864 in mated females (discussed below).

We also drove expression of these transgenes in females using a ubiquitous driver (*tubulin-GAL4*). These females did not carry the CG11864 null allele because combination of the CG11864 null chromosome 2L with the *tubulin-GAL4* chromosome results in lethality. CG11864 expressed in females from the transgenes has a higher apparent molecular weight than the endogenous CG11864 protein in males (Figure 3.5C). CG11864 expressed in females from the wild-type and Q155E transgenes appears on a Western blot as a doublet. Both bands of this doublet are larger than endogenous CG11864 in male AG (Figure 3.5C, lanes 3 and 7). The lower molecular weight band of this doublet may correspond to the cleaved, activated form of CG11864. Consistent with this, only the higher molecular weight band is present in females expressing the cleavage mutant form of CG11864 (Figure 3.5C, lane 5). RTs of the sibling balancer controls of these females have no cross-reactivity with the CG11864 antibody within the size range of the doublet (Figure 3.5C, lanes 4, 6, and 8), indicating that the doublet seen in the transgene-expressing females is in fact CG11864.

The higher apparent molecular weight of CG11864 in females expressing CG11864 transgenes may indicate that this protein is post-translationally modified improperly in female tissues. To determine if the gel mobility difference might be a result of a glycosylation difference between male AG and female tissues, we treated CS male AG and virgin female RTs expressing the cleavage mutant CG11864 transgene with the enzyme PNGaseF (Figure 3.5D). PNGaseF removes N-linked glycosylation marks from proteins [222]. CG11864 from male AG treated

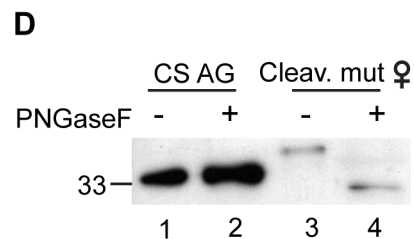
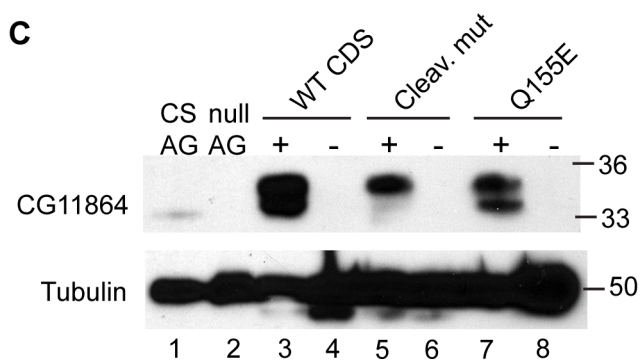
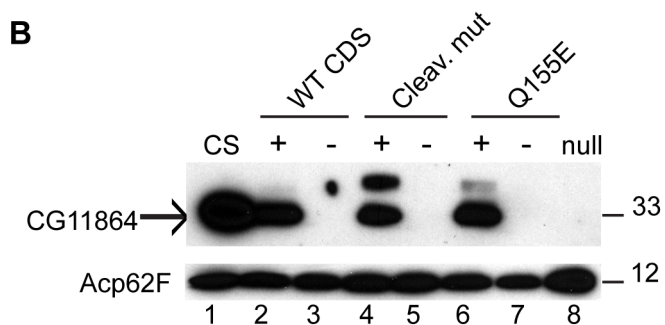
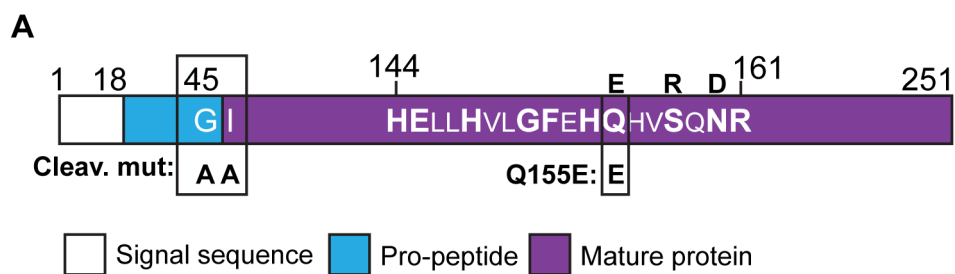
with PNGaseF did not change in apparent molecular weight on a Western blot relative to untreated AG, indicating that CG11864 does not normally carry N-linked glycosylation marks (Figure 3.5D, lanes 1 and 2). In contrast, cleavage mutant CG11864 expressed in virgin female RTs did shift in size after PNGaseF treatment, to an apparent molecular weight of a few kiloDaltons less than CG11864 in CS males. This shift indicates two things: first, CG11864 expressed in females is N-glycosylated, unlike CG11864 expressed in male AG. Second, CG11864 in male AG is post-translationally modified, but not by N-glycosylation.

**Figure 3.5: UAS-CG11864 expression in CG11864 null males and in virgin females.**

**(A)** Model of the CG11864 protein. White: the predicted secretion signal sequence; Blue: the predicted pro-peptide sequence; Purple: the predicted mature sequence. Numbers indicate amino acid starting from Met1. The extended astacin-family catalytic site sequence is at amino acids 144-161; conserved residues are in bold and large font; for those that differ from the consensus, the consensus site is shown above the CG11864 sequence. Mutations made in UAS/Gal4 driven constructs are shown below the protein model.

**(B)** Western blot probed with anti-CG11864 or Acp62F antibody (latter used as a loading control). Each lane contains AG from 1 male. Lane 1: AG from CS male. Lanes 2-3: AG from males carrying the wildtype UAS-CG11864 transgene at the *Attp40* genomic locus. Lanes 4-5: AG from males carrying the cleavage mutant CG11864 transgene at the *Attp40* locus. Lanes 6-7: AG from males carrying the Q155E CG11864 transgene at the *Attp40* locus. Lane 8: AG from CG11864 null male. (+): transgene driven with CG11864 null; *paired-Gal4* driver. (-): Balancer siblings of + males (these do not carry the driver).

**(C)** Western blot probed with anti-CG11864 and alpha-Tubulin (loading control) antibodies. Lane 1: 1 pair AG from CS male. Lane 2: 1 pair AG from CG11864 null male. Lanes 3-8: Each contain 10 virgin female RTs from females expressing WT CG11864 (Lane 3), Cleavage mutant CG11864 (Lane 5), or Q155E CG11864 (Lane 7) driven by *Tubulin-Gal4*, or their balancer sibling controls (Lanes 4, 6, and 8). **(D)** Western blot probed with anti-CG11864 antibody. PNGaseF is used to remove N-linked glycosylation marks from proteins [222] Lanes 1-2: 1 pair AG each from CS males. Lanes 3-4: 5 virgin RTs from females expressing the cleavage mutant form of CG11864. (-): Untreated samples. (+): Samples treated with PNGase F.



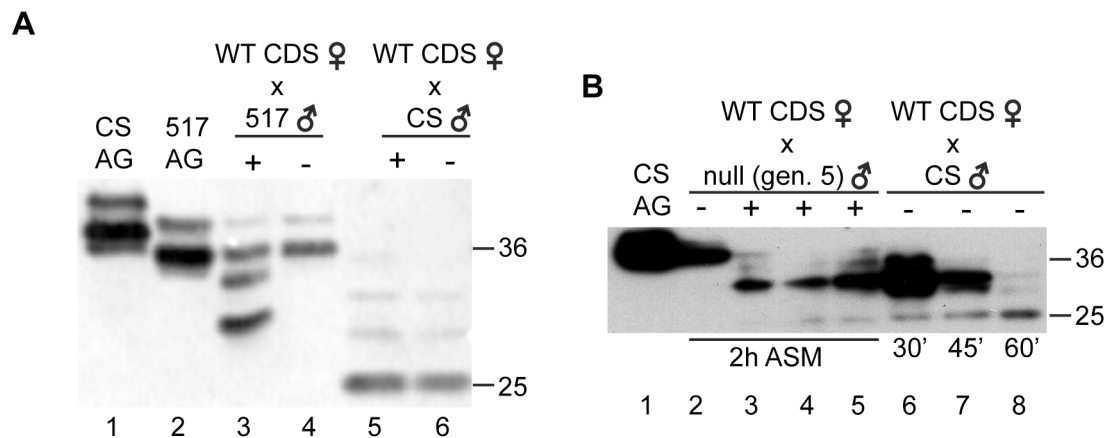
### ***Female-expressed CG11864 rescues processing defect of mating with CG11864 null male***

We next tested the wild-type CG11864 produced from a transgene for its ability to cleave ovulin when ectopically expressed in the female. CG11864 is normally cleaved, and presumably thereby activated, during mating while still in the male reproductive tract. If factors specific to the ejaculatory duct and/or bulb (that are not transferred to and/or active in females) are required for CG11864 activation, it should not be functional when ectopically expressed in females. In addition, the non-native N-glycosylation of ectopically expressed CG11864 in females suggests that there are sex-specific post-translational modifications of CG11864, and this may have consequences for CG11864 catalytic activity.

I tested the ability of ectopically expressed CG11864 in females to rescue the CG11864 null mutation in males using two independent transgenic lines. First, we used females expressing CG11864 from a randomly inserted transgene (UAS-CG11864 ; *tubulin-Gal4*)[65]. These females express CG11864 identically to the females used in Figure 3.5 (data not shown). We mated these females (and their sibling controls, which do not express CG11864) to DGRP-517 males. These males transfer ovulin and other Sfps, but not CG11864, to females. At 1 hour ASM, females expressing the transgene contained partially cleaved ovulin (Figure 3.6A, lane 3), in contrast to wild-type females mated to DGRP-517 males, which only contained intact ovulin (Figure 3.6A, lane 4). Similar results were obtained with females expressing CG11864 from a transgene inserted into the *Attp40* locus (see Figure 3.5C). At 2 hours ASM, the 25kDa ovulin cleavage product (from the final cleavage event) is present in transgenic females mated to CG11864 null males (Figure 3.6B, lanes 3-5). However, there is still more of the full-length ovulin and its intermediate cleavage products in transgenic females mated to CG11864 null males than in these females mated to CS males at earlier times ASM (Figure 3.6B, lanes 6-8).

Sibling control females (not expressing CG11864) mated to CG11864 null males contain only full-length ovulin at this timepoint (Figure 3.6B, lane 2). 2 hours ASM is the latest timepoint at which ovulin can be reliably detected in RTs of mated females (personal observation; data not shown).

These data show that females' ectopically-expressed CG11864 can compensate for the lack of CG11864 from the male. In females that had mated to CG11864 null males, the female-expressed CG11864 is able to cleave both ovulin and Acp36DE, albeit more slowly than the male-expressed protease (Figure 3.6A, lanes 5 and 6). Thus, CG11864 does not need to be synthesized in, or travel through, the male reproductive tract to be active.



**Figure 3.6: CG11864 expressed in females can process ovulin. (A)** Western blot probed with anti-ovulin antibodies. Lanes 1-2: 1 pair AG each from indicated males. Lanes 3-4: 4 female RTs 1 hour ASM to DGRP-517 males. Lanes 5-6: 4 female RTs 1 hour ASM to CS males. (+): Females expressing a randomly inserted CG11864 WT CDS transgene. (-): Sibling control females. **(B)** Western blot probed with anti-ovulin antibody. Lane 1: 1 pair AG from CS male. Lanes 2-5: 10 female RTs 2 hours ASM after mating to CG11864 null males (backcrossed to *w<sup>1118</sup>* for 5 generations). Lanes 6-8: 4 female RTs after mating to CS males at 30, 45, and 60 minutes ASM. (+): Females expressing WT CG11864 transgene from Attp40 genomic locus. (-): Sibling control females.

Previous work has shown that CG11864 is normally cleaved within the male ejaculatory duct and/or bulb during ejaculation by the trypsin-like Sfp seminease [63, 81]. Our results here indicate that transit through the male reproductive tract is not required for this activation. On the other hand, CG11864 expressed in females is not as efficient in cleaving ovulin or Acp36DE as is endogenous CG11864 provided by males. This may be for several reasons. First, females express an improperly modified form of CG11864, which may impair its activity. Second, females may express CG11864 at a lower level than they would receive from males or may not secrete as much protein into the lumen of the reproductive tract. Finally, seminease may not be as efficient at activating CG11864 in females as it is in males, though this seems less likely as CG11864 continues to be cleaved after deposition into the female in normal matings [81].

### ***The female reproductive tract environment is not necessary for the first two ovulin cleavages***

The contribution of the female reproductive tract environment to ovulin processing is not well understood. Ovulin cleavage products are never detected while ovulin is still in the male reproductive tract [106]. Conversely, females themselves cannot cleave ovulin [106] because male-derived Sfps (i.e., CG11864) are required for ovulin cleavage, as well as for the cleavage of Acp36DE [81].

We asked whether females contribute functional proteins (e.g., proteases), or simply the optimal environment (e.g., correct pH or ion concentrations) or sufficient time, to the processing of either ovulin or Acp36DE, by collecting male ejaculates in the absence of females. We used males ectopically expressing the temperature-sensitive neuronal activator dTrpAI in neurons required for courtship and mating behaviors (male genotype: *dTrpAI/+; fru-GAL4/+*). Courtship and mating neuronal pathways are activated in these males by exposure to high temperatures

[223]. These males will ejaculate after a few minutes at 29°C, in the absence of females (Figure 3.7A). The ejaculate is often present as a discrete, semi-solid mass near the external genitalia or on the male's abdomen (Figure 3.7A, top panel). However, it is not always fully ejaculated at the time of collection, and males can sometimes ejaculate into a droplet of buffer. The first portion of the ejaculate appears as a solid mass and is followed by a more liquid portion containing sperm and, presumably, Sfps (Figure 3.7A, bottom panel). Due to this variation, it is difficult to know whether we successfully collected the entire ejaculate from each male.

We collected ejaculates from several males to test for the presence of ovulin and Acp36DE cleavage products. As a control for the *ex vivo* processing rate of ovulin, we also dissected female RTs at 10 minutes ASM (before ovulin cleavage products are normally detected) and incubated female RT protein extracts in the same buffer used for ejaculate collections. By 1 hour after dissection, ovulin cleavage products can be detected in the *ex vivo* samples (Figure 3.7B, lane 4). This processing is much slower than that seen in females dissected at 1 hour ASM, possibly due to sub-optimal buffer conditions or to the increase in volume relative to the female reproductive tract.

Ejaculates were incubated for various lengths of times in two different buffers. In buffer A, a band with the same apparent molecular weight as ovulin CP2-C, approximately 30kDa, was visible in some samples (Figure 3.7B, lane 6; refer to Figure 3.1 for cleavage product sizes), but not in others (Figure 3.7B, lanes 7 and 8).

We also looked for evidence of Acp36DE cleavage. Acp36DE is cleaved from a 122kDa precursor to two cleavage products of 68 and 50kDa, respectively; the 50kDa product is not visible in the blot in Figure 3.7B. The 68kDa cleavage product of Acp36DE appeared in all male ejaculate samples, and possibly increased in abundance over time (Figure 3.7B, lanes 5-8),



though this is not absolutely clear since the Westerns are not quantitative.

In buffer B, we were only able to detect ovulin cleaved to the second cleavage product (CP2-C) in the *ex vivo* female RT samples after 1 hour of incubation (Figure 3.7C, lane 4; compare to Figure 3.7B, lane 4). This may have been a result of lower total protein loading in these samples, relative to those used in the previous experiment.

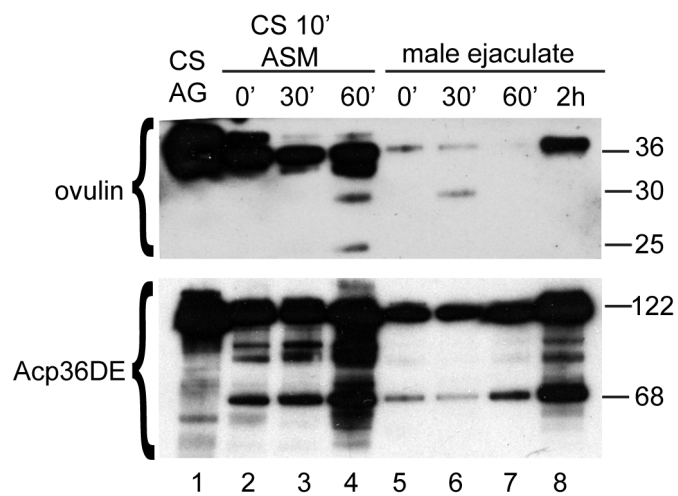
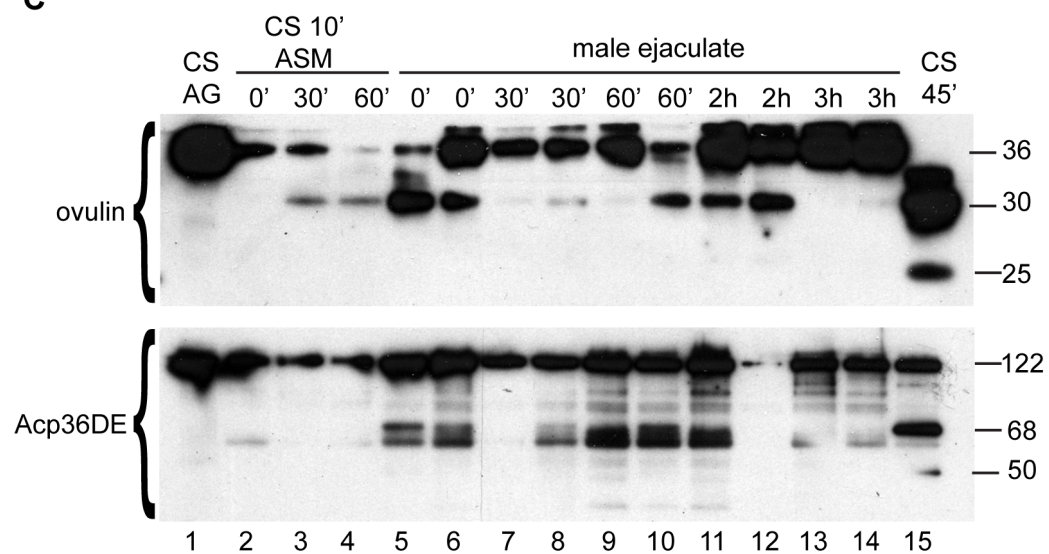
The majority of male ejaculates in buffer B appeared to contain ovulin processed to the same extent as in the *ex vivo* female samples (Figure 3.7C, lanes 5-14), though the level of processing did not correlate with time spent in incubation; CP2-C was present in samples incubated for 2 hours (Figure 3.7C, lanes 11-12), but not in those incubated for 3 hours (Figure 3.7C, lanes 13-14). A potential cleavage product was also seen in male ejaculate samples that were not incubated for additional time after collection (total time before addition of SDS buffer in these samples is approximately 10 minutes after ejaculation) (Figure 3.7C, lanes 5 and 6). This product has the same apparent molecular weight as CP2-C from a wild-type mating (Figure 3.7C, lane 15), suggesting that it is indicative of ovulin cleavage in the ejaculate samples.

Acp36DE processing was not detected in the *ex vivo* samples in buffer B (Figure 3.7C, bottom panel, lanes 2-4). In contrast, the 68kDa cleavage product of Acp36DE was detectable in many of the ejaculate samples (Figure 3.7C). Again, this may be a function of uneven protein loading. A suitable loading control protein could not be identified for these experiments. A standard loading control, such as Tubulin could not be used because the samples contain only extracellular proteins and may not have equal numbers of sperm. Other Sfps are not suitable as loading controls because they may be unequally distributed within the ejaculate, and thus may not reflect the total protein content if certain portions of the ejaculate are more prone to loss during collection.

At least in one buffer tested, RTs of females dissected at 10 minutes ASM and incubated at room temperature *ex vivo* were able to process both ovulin and Acp36DE more quickly than ejaculates collected from males. However, the slow rate of processing in *ex vivo* female RTs compared to *in vivo* processing suggests that more suitable buffer and/or temperature conditions may exist that would allow ovulin to be more fully processed in the ejaculate samples.

**Figure 3.7: Ovulin can be partially cleaved in ejaculates of males in the absence of females.**

- (A) *UAS-dTrpAI/+;fru<sup>GAL4(B)</sup>/+* males after heat activation. Seminal fluid is collected from males after ejaculation with forceps in a drop of buffer. Ejaculate is shown in top panel (white arrowhead). Bottom panel shows ejaculate from a single male, separated from the male's abdomen. A semi-solid mass emerges first (white arrowhead), followed by the more liquid, apparently sperm-containing portion (curly bracket).
- (B) Western blot probed with anti-ovulin and Acp36DE antibodies. Lane 1: 1 pair AG from CS male. Lanes 2-4: 3 female RTs dissected at 10 minutes ASM to CS male and incubated in buffer A for indicated time. Lanes 5-8: Ejaculates collected from 3 males in each lane and incubated in buffer A for the indicated time (following heat activation and time spent in buffer during collection).
- (C) Western blots probed as in (B); samples were collected as in (B), except incubated in buffer B. Lane 15 contains 3 female RTs 45 minutes after mating to CS males for comparison.

**A****B****C**

These experiments indicated that initiation of ovulin and Acp36DE processing does not require any contribution from the female. However, the *in vitro* processing in collected ejaculates was inefficient and not consistently reproducible. Furthermore, we never observed ovulin cleavage to proceed past the second cleavage event in ejaculates incubated *in vitro*. Thus, female proteins (such as additional proteases) may be required for the third cleavage event based on these data. The female may contribute protein co-factors that enhance the efficiency of CG11864, or they may simply provide the optimal pH or ionic concentration for CG11864 activity. Further experimentation is required to determine which female factors are necessary, if any, for ovulin and Acp36DE processing.

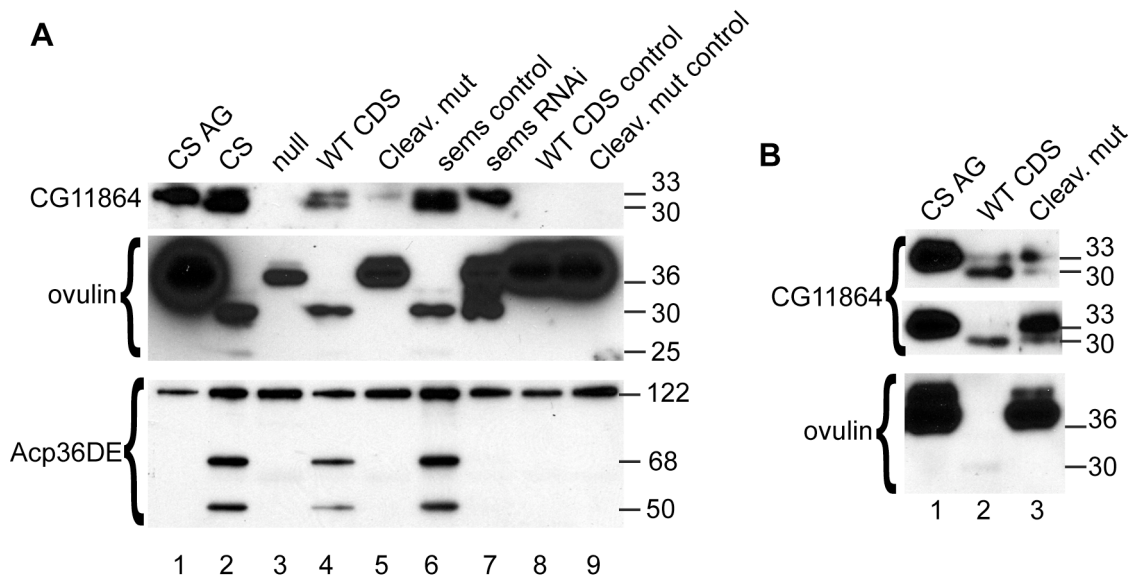
***CG11864 expressed from the cleavage mutant transgene fails to cleave ovulin or Acp36DE***

We recently reported that the trypsin-like serine protease seminepsin is required for the predicted activation cleavage of CG11864 [63]. CG11864 is synthesized as a predicted inactive zymogen [81]; prediction of its cleavage site was based on sequence threading to other astacin-family metalloproteases [81]. Removal of the pro-peptide region of CG11864, as in many other proteases [198], is predicted to be required for its activation [63, 81].

We tested the prediction that driving expression of CG11864 with mutations at this predicted cleavage site—in the CG11864 null background—in male accessory glands would generate an uncleavable form of CG11864 and loss of both ovulin and Acp36DE processing in the mated female. We will refer to the transgene used to express the predicted uncleavable form of CG11864 as the cleavage mutant transgene.

We drove expression of the wild-type CG11864 transgene and the cleavage mutant transgene (see Figure 3.5) in CG11864 null male AG and mated these males to CS females

(Figure 3.8). At 45 minutes ASM, females mated to males expressing the wild-type transgene contained processed ovulin (CP1-C and CP2-C at 33kDa and 30kDa, respectively) and Acp36DE (68- and 50kDa bands) (Figure 3.8A, lane 4). Ovulin from females mated to males expressing the wild-type CG11864 transgene were indistinguishable from females mated to CS males (Figure 3.8A, lane 2). Thus, the wild-type CG11864 transgene was able to fully complement the ovulin and Acp36DE processing defects in CG11864 null males (Figure 3.8A, lane 3).



**Figure 3.8: Predicted activational cleavage site of CG11864 is required for ovulin and Acp36DE processing. (A)** Western blots probed with anti-CG11864 (top), ovulin (middle), and Acp36DE (bottom) antibodies. Acp36DE blot is from a separate gel. Lane 1: 1 pair AG from CS male. Lanes 2-9: Mated female RTs (2 each well) at 45 minutes ASM to indicated male. Sems control: female mated to balancer sibling of sems RNAi male. Sems RNAi: female mated to male knocked down for seminal. Lanes 8-9: females mated to balancer siblings of males used in lanes 4-5. **(B)** Western blot probed with anti-CG11864 antibody (both panels from the same gel) and anti-ovulin antibody (also from the same gel; ovulin samples correspond to the same lanes as in the lower CG11864 panel). Lane 1: 1 pair AG from CS male. Lanes 2-3: 2 female RTs at 1 hour ASM. Lane 2: mated to male expressing WT CG11864 in null background. Lane 3: mated to male expressing cleavage mutant form of CG11864 in null background.

In contrast, only full-length ovulin and Acp36DE were detected in females mated to males expressing the cleavage mutant form of CG11864 (Figure 3.8A, lane 5); ovulin gel mobility patterns in these females were indistinguishable from those mated to CG11864 null males (Figure 3.8A, lane 3). The ovulin cleavage defect was more severe than that seen with *seminase* or CG11864 RNAi males (Figure 3.8A, lane 7) in which ovulin is cleaved to the second cleavage product (CP2-C is at 30kDa).

The cleavage mutant form of CG11864 did not remain fully intact in females mated to cleavage mutant males (easier to see in the samples in Figure 3.8B). Both full-length CG11864 and a band of approximately the same molecular weight as cleaved CG11864 were present in the RTs of females mated to cleavage mutant males at 1hour ASM. Ovulin at this time was not processed (Figure 3.8B, lane 3). However, full-length CG11864 is still relatively more abundant than the cleaved form in females mated to cleavage mutant males (Figure 3.8B, lane 2). Two representative samples are shown in Figure 3.8B. For comparison, cleaved CG11864 is undetectable in females mated to *seminase* RNAi males (Figure 3.8A, lane 7 and [63]), even though these females do slowly process ovulin.

One possible explanation for the apparently cleaved but inactive CG11864 is that CG11864 might have two cleavage sites within a few amino acids of each other and must be cleaved at both sites for activation. By mutating the I46 cleavage site, we may have blocked the secondary cleavage event, but not the first. An activation mechanism like this exists for the canonical *Astacus astacus* astacin protease [221]. Astacin is cleaved first by a trypsin protease a few amino acids N-terminal to the actual pro-peptide cleavage site. Following this tryptic cleavage event, the remaining residues are removed from the pro-peptide by an intramolecular cleavage of astacin itself, producing the active protease [221]. We suggest that a similar

mechanism may be used to activate CG11864.

***CG11864 expressed from the Q155E transgene cleaves ovulin and Acp36DE more slowly than wild-type transgenic CG11864***

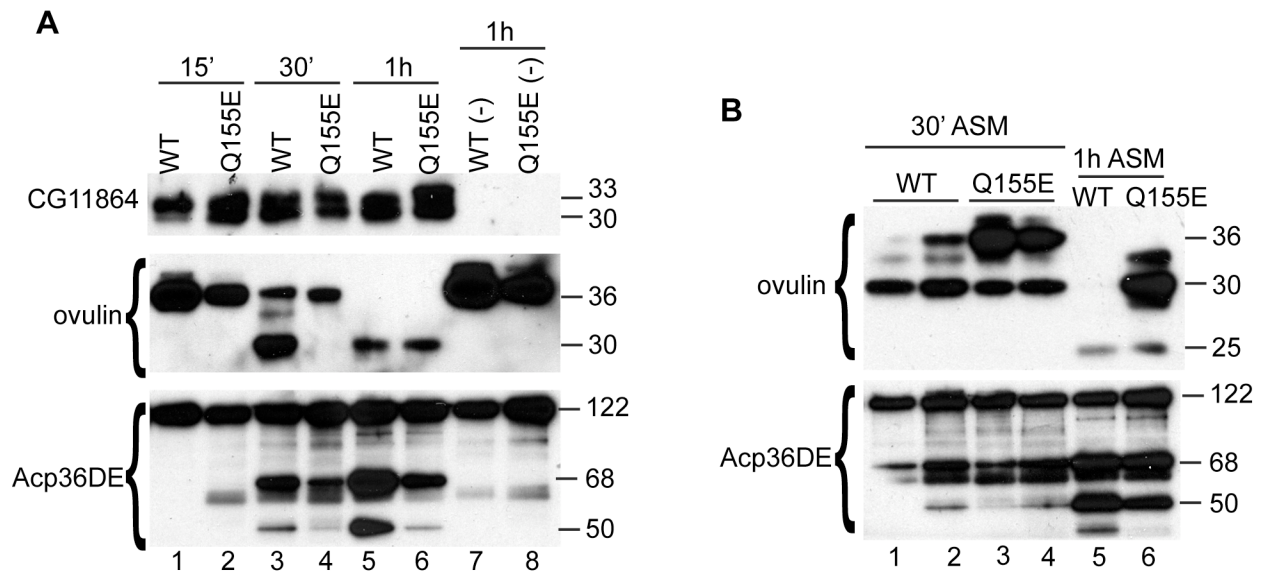
We tested the role of a conserved catalytic residue (Q155) that differs from the homologous residue of all other astacin proteins. Q155 (astacin E103) is important for coordinating the catalytic zinc ion [224]; mutating this residue to Ala results in complete loss of catalytic activity [221]. This residue is nearly universally conserved as Glu in astacin-family proteins [224]. When the corresponding residue in astacin, Glu103, is mutated to Gln, astacin becomes extremely heat labile [221], though catalytic activity at physiological temperatures is not significantly affected.

CS virgin females were mated to males null for endogenous CG11864 and that expressed either the WT CG11864 transgene or the Q155E transgene. CG11864 was cleaved similarly in females mated to both types of males (Figure 3.9A, CG11864 blot). At 15 minutes ASM, neither ovulin nor Acp36DE was cleaved in females mated to either type of male (Figure 3.9A, lanes 1 and 2). By 30 minutes ASM, ovulin from males expressing the WT transgene is cleaved to the second cleavage product, whereas ovulin from Q155E males is still full-length (Figure 3.9A, lanes 3 and 4).

In a separate experiment comparing females mated to both types of male at 30 minutes ASM, ovulin from Q155E males was partially processed, though processing was still slower than in WT transgenic males (Figure 3.9B, lanes 1 and 2). In some experiments, ovulin processing in females mated to Q155E males had caught up to that in WT-mated females by 1 hour ASM (Figure 3.9A, lanes 5 and 6), while in others it was delayed (Figure 3.9B, lanes 5 and 6).



The effect of the Q155E mutation on Acp36DE processing was subtler. At 30 minutes and 1 hour ASM, both Acp36DE cleavage products could be detected in females mated to both types of males, however the abundance of the cleavage products was higher in females mated to WT transgenic males (Figure 3.9; the result is clearer at 30 minutes ASM). An additional cleavage or degradation product of less than 50kDa in Acp36DE is often visible at times after 1 hour ASM. In Figure 3.9B, this band is clearly more abundant at 1 hour ASM in females mated to WT transgenic males than in females mated to Q155E males (lanes 5 and 6).



**Figure 3.9: Conserved catalytic site mutation in CG11864 results in slower ovulin processing.** (A) Western blot probed with anti-CG11864 (top), ovulin (middle), and Acp36DE (bottom) antibodies. Each lane contains 2 female RTs frozen at the time indicated (ASM) to male expressing WT CG11864 or Q155E CG11864 transgene in CG11864 null background (Lanes 1-6) or their sibling controls (Lanes 7-8). (B) Additional replicates of the 30 minutes and 1 hour ASM samples from (A).

Our results indicate that CG11864 has adapted to use the uncharged residue Gln in place of the canonical, negatively charged Glu. Substitution of this residue with Glu results in lowered efficiency of CG11864-mediated cleavage of its substrates. The effect of the Q155E substitution on processing of ovulin and Acp36DE varied among experiments. The variation in results may be due to differences in temperature, substrate protein abundance, or individual interactions between males and females during mating. However, the overall trend was in the direction of slower CG11864-mediated processing with E155 compared to Q155. These results indicate there is something unique about either the female reproductive tract environment or the surrounding residues in CG11864 that renders a Gln in this position beneficial in contrast to Glu.

#### ***CG11864 null males have no obvious fertility defects***

We tested CG11864 null males for their ability to induce ovulation and egg-laying in mated females (see Methods for description of experiments). CG11864 null males were able to induce rates of ovulation and levels of egg-laying similar to those of their controls in all experiments (Table 3.2).

**Table 3.2: Fertility assays performed with CG11864 null males.** Experiments were performed with CG11864 null males compared to their controls, each backcrossed 10 generations into a standard laboratory fly stock ( $w^{1118}$ ). See Materials and Methods for descriptions of experiments.

Experiment	N P{EP}CG5861	N CG11864 null	Statistic (df)	P value
Ovulation assay	84	74	t = -0.81 (154.55)	0.42
24 hour egg count	129	131	t = -0.33 (257.96)	0.74
10 day egg count	101	101	t = -1.14 (193.55)	0.26
5 day egg count (2 day old females)	75	64	t = 0.55 (113.4)	0.584
3 day egg count (10-12 day old females)	42	33	t = -1.13 (71.72)	0.264

In addition to a standard 10-day egg-laying assay [105], we also tested whether there was an effect of CG11864 on egg-laying that varied with female age. Females of different ages than the 3-5 days post-eclosion females used in the 10-day egg-laying assay may respond differently to the absence of CG11864 because of the availability of different numbers of mature oocytes, hormone levels, or other unknown factors affected by age.

In one experiment, we mated females that were about 48 hours post-eclosion; female *Drosophila melanogaster* are considered sexually mature at 3 days post-eclosion. No fertility defects were associated with the CG11864 null allele in these experiments (results for total egg counts are shown in Table 3.1, though individual days' egg counts were also analyzed, with similar results).

We also performed an experiment with much older females. Unpublished results from T. Chapman indicate that females lose the capacity to respond to sex peptide as they age. Sex peptide has a large effect on egg-laying and may overwhelm any effect of CG11864. We mated females between 10-12 days post-eclosion to CG11864 null males or controls, but did not see any significant difference in the numbers of eggs laid (data from 3 day totals shown in Table 3.2, though all individual days tested were also non-significant).

These results were not entirely unexpected, as deletion of ovulin only has a modest effect on egg-laying in the first 24 hours ASM [105] and Acp36DE deletion does not affect egg-laying [6]. In addition, full-length ovulin can stimulate ovulation, though there is evidence that the cleavage products may have a somewhat stronger stimulatory effect [108].

Our data are also consistent with the lack of any observable fertility defect reported for CG11864 RNAi males [81]. CG11864 may only have a measurable fertility effect under field conditions or other specific, non-standard laboratory conditions, which have yet to be tested.

These may include starvation conditions, rearing and/or mating at different temperatures, varying female and/or male mating status, or varying the social context (e.g., mating in the presence of a second male). Another possibility is that proteolytic cleavage of ovulin, Acp36DE, and potentially other substrates, does not have important functional consequences in this species, though it seems unlikely that *D. melanogaster* would preserve a functional CG11864 gene if there were no consequences to its loss. It may be that we have simply not assayed the sensitive phenotype. Future experiments to determine the fate of ovulin and Acp36DE cleavage products, possibly by expressing these products in male accessory glands in the absence of endogenous full-length protein, may help to understand their roles in post-mating processes.

## DISCUSSION

We report here a naturally-derived null allele of CG11864. The null allele is caused by a *blood* transposable element inserted into the first exon of the CG11864 gene. The *blood* retrotransposons are a family of transposable elements containing long terminal repeats (LTR element) with a canonical length of 7,410bp [225]. Females mated to males homozygous for this disrupted allele of CG11864 do not process either ovulin or Acp36DE. These results indicate that CG11864 is necessary for all ovulin and Acp36DE cleavage events. Previous results using RNAi knockdown of CG11864 [81] that showed only partial loss of processing of these two Sfps must have reflected residual CG11864 presence.

We have also determined that the predicted pro-peptide cleavage site of CG11864 is required for its activity, further supporting the hypothesis that the CG11864 cleavage seen in the male ejaculatory duct and/or bulb during mating is activational. However, since not all CG11864

cleavage was lost in our experiments, we cannot be certain that our mutation fully disrupted cleavage at this site. Our data suggest that secondary cleavage sites may exist near the activational cleavage site tested here. Cleavage at these sites does not appear to activate CG11864. The role of any potential secondary cleavage sites should be investigated further.

Previous work demonstrated that while ovulin is not processed in male accessory glands, neither can it be processed alone by the female [106]. Later studies found that two male proteases, CG11864 [81] and seminare [63] are required for ovulin processing, though they are only apparently active in the female reproductive tract. The current study has further dissected the respective roles of males and females in ovulin (and Acp36DE) processing. Ectopic expression of CG11864 in females allowed for ovulin processing after mating to CG11864 null males. Thus, the female provides an environment suitable for CG11864 activational cleavage (even though that cleavage is normally initiated in the male [63, 81]). The functions of Sfps in mated females often appear to conflict with the interests of the female, while other Sfp roles indicate cooperation between the sexes [226]. Based on the results of this and previous studies, ovulin processing would seem to fall into the latter category.

We also observed that ectopic expression of CG11864 in virgin females resulted in two isoforms of CG11864, visible as two bands on a gel. In females expressing a cleavage mutant form of CG11864, only the larger band is visible. These results suggest that CG11864 may be cleaved, and possibly activated in females prior to receiving other male proteins, including seminare. However, whether female-expressed CG11864 is active in the absence of seminare still needs to be determined.

We also investigated whether females contribute any necessary factors for ovulin cleavage when ovulin, CG11864, and seminare are all present. Ejaculate samples collected from

males in the absence of females were able to partially process ovulin. Acp36DE cleavage was also observed in these samples. However, processing was much slower than in either *in vivo* or *ex vivo* female reproductive tracts. This result indicates that the female provides factors (such as protein co-factors, or optimal pH or ionic concentrations) necessary for efficient Sfp processing.

The results of both the female ectopic expression experiments and male ejaculate collection experiments underscore the importance of characterizing the conditions within the female tract. To date, no measurements have been done to determine the pH or important ion concentrations (e.g., zinc—CG11864 is predicted zinc metalloprotease) within the female uterus. These experiments would be invaluable to studies of *Drosophila* Sfp function in general, and would help determine what factors the female provides specifically for ovulin and Acp36DE processing.

With this study, we have gained further insight into the regulation of a proteolytic pathway in *Drosophila* seminal fluid. This apparent cascade begins with the activation of seminase during transit through the male reproductive tract at mating, leading to the activation of CG11864 and processing of ovulin and Acp36DE [63]. We have determined here that both male and female components are important for this pathway, even though males appear to provide the necessary proteases for at least partial processing to occur. Our results suggest that conditions of the mated female reproductive tract together with a CG11864-specific catalytic site substitution (i.e., Q155) have been fine-tuned for an optimal rate of ovulin processing. While the functional consequences of either ovulin or Acp36DE cleavage remain unclear, the demonstration of both male and female cooperation in this proteolytic pathway makes it an interesting model system for future studies of the regulation and evolution of seminal fluid protease cascades.

## METHODS

### *Flies*

We reared flies under standard laboratory conditions on yeast/glucose media at  $23\pm 2^{\circ}\text{C}$  on a 12:12 light/dark cycle. We are grateful to T. Mackay (NCSU) for the DGRP lines, including 517, from which we derived the CG11864 null stock using the crossing scheme outlined in the Supplementary Methods (part 1). We used Bloomington stock 27436 ( $P\{w[mC]EP\}CG5861[G18171]$ ) to link the CG11864 null allele to a *mini-white* allele. Failure of PCR amplification using CG11864 primers was used to test for retention of the CG11864 null allele. We backcrossed to the CG11864 null line into  $w^{1118}$  for 10 generations. In parallel, we backcrossed stock 27436 into  $w^{1118}$  for 10 generations in order to use it as the control in fertility experiments (“control males”). “CG11864 null” or “null” males used in experiments were, unless otherwise noted:  $w^{1118}; \{blood\}CG11864, P\{EP\}CG5861$  backcrossed 10 generations to  $w^{1118}$  (line C). We used Bloomington stock 7831 ( $w^{1118}; Df(2L)Exel7063/CyO$ ) to test if the deficiency could complement the CG11864 defect in DGRP-517. We used seminasal control and RNAi males as described in [63]. The *paired*<sup>GAL4</sup> driver line was purchased from the Bloomington stock center (stock number 1947).

We generated the females used in the experiment in Figure 3.6A (ectopic expression of CG11864) by crossing the previously described *pUAST-CG11864* line [65] to a ubiquitous driver line (*tubulin-GAL4 / TM3*) obtained from the Bloomington stock center (stock #5138). We are grateful to B. Baker (Janelia Farms) for flies used in ejaculate collection experiments. These include *UAS-dTrpAI* [227] and *UAS-mCD8-Gfp; fru-GAL4(B) / MKRS* [228]. Males carrying both the *UAS-dTrpAI* and *fru-GAL4(B)* (derived from the cross between these two lines) express

courtship and mating behavior after exposure to heat (29°C) for a few minutes due to activation of the temperature-sensitive neuronal activator dTrpAI [223].

### ***Ovulin Gene Sequencing***

The ovulin gene was sequenced from genomic DNA of line DGRP-517 to determine the nature of the ovulin protein truncation. Sanger sequencing was performed by the Cornell BioResources core facility. We used the following primers for PCR and Sanger sequencing:

OvulinForward: AGC AGA ACG AAA CAA GAG CC

OvulinReverse: AGA ATC CAA CCA GCA CTT GG

### ***Cloning of CG11864 Constructs into the pBI-UAS-CG Vector***

We used the Gateway (Invitrogen) system-compatible pBI-UAS-CG vector [229] to clone the Canton-S coding region of CG11864. This vector is compatible with the PhiC31 site-specific integrase system in *Drosophila* [230-232]. In addition to the wildtype CG11864 sequence, we cloned two constructs with mutations made using the QuikChange Lightning kit (Agilent): G45A, I46A (cleavage mutant) and Q155E. Genetics Services, Inc. (Cambridge, MA) injected all three constructs into *y,w/++* embryos expressing the PhiC31 integrase and carrying the Attp40 site on chromosome 2L at cytogenic location 25C7 [233]. We crossed individual injected flies to *w<sup>1118</sup>* and selected *w+* offspring. Individual transgenic flies were crossed again to *w<sup>1118</sup>*. We established homozygous transgenic lines from these crosses.

### ***Transgene expression***

To express CG11864 in females, we crossed females of the CG11864 wildtype transgenic



line to *tubulin-Gal4 / TM3* driver males. We sorted female offspring by the TM3 visible marker Stubble (*Sb*): *Sb*<sup>+</sup> females express Gal4 (and thus, the transgene); *Sb* females (balancer controls) do not.

We crossed each transgenic line into the CG11864 null background for male expression experiments. The Attp40 landing site is greater than 50 map units from the CG11864 locus, so we were able to easily obtain recombinant flies. Briefly, CG11864 null flies were crossed to a transgenic strain. Resultant females were crossed to males of a line with two different second chromosome balancers (*w;CyO/Gla*). We identified recombinant flies by their greater expression of the *mini-white* transgene (a hypomorphic allele of the *white* gene linked to the CG11864 transgenes and CG11864 null allele). We then crossed Curly (*CyO*) recombinant flies to each other to generate homozygous stocks.

To generate males expressing each of the transgenes in the CG11864 null background, we crossed females from the CG11864 null transgenic lines to the *CG11864null; paired-GAL4 / TM3* driver males derived from the crosses in the Supplementary Methods (part 2). The *paired-GAL4* driver drives expression of the transgene in the male accessory glands [220]. We identified male progeny from the cross that carried both the transgene and the driver by the visible marker *Sb*<sup>+</sup>.

### ***Deglycosylation of female-expressed cleavage mutant CG11864***

To determine whether female-expressed CG11864 was improperly post-translationally modified by addition of N-linked glycosylation marks, we treated the expressed proteins with the enzyme PNGaseF (NEB). Male AG (2) or virgin female RTs (10) were dissected into and ground in the supplied glycoprotein denaturing buffer. Each sample was then split in two. We boiled the

treated sample with SDS sample buffer for Western blot analysis. We added 1X deglycosylation buffer (supplied with the enzyme), NP-40 to a final concentration of 1%, and 1,000U PNGaseF to the treated samples. Treated samples were incubated at 37°C for 1 hour. We then prepared the treated samples for Western blot analysis by boiling with SDS sample buffer.

### ***Observed Matings***

We conducted all matings by placing single pairs of 3-6 day old virgin females and males in glass vials. Each vial contained a small piece of moistened filter paper. Unless otherwise noted, all females used for Western blot analysis were from the Canton-S (wildtype) strain. All females used for fertility assays were from an 8-generation inbred population of the  $w^{1118}$  (white-eyed, wildtype) strain. This strain was used to reduce genetic variation among females. We noted the time at which each mating began; we discarded mating pairs with unusually short (<15 minutes) copulation durations. For Western blot samples, we individually froze mated females in liquid nitrogen at the time after the start of mating (ASM) specified for each experiment. For fertility experiments, females were placed individually in plastic vials with standard yeast/glucose media immediately after the end of mating. For the ovulation assay experiment, females were then placed on ice in 0.5-mL Eppendorf tubes at 3 hours ASM, until dissection.

### ***Sample preparation and Western blot analysis***

We dissected whole male accessory glands (AG) or female lower reproductive tracts (uterus and sperm storage organs, RT) in PBS under a dissecting microscope. We placed tissues were approximately 10μL of PBS in a 0.5-mL Eppendorf tube and briefly ground them with a pestle and added 10μL of 2x SDS sample buffer containing 10% 2-mercaptoethanol to the

samples and boiled them for 5 minutes.

We separated samples using SDS/PAGE. Most gels were 12% polyacrylamide except for those in Figures 3.2A, B, and D, 4D, and Figure 3.6B which were 5-15% gradient gels. Western blotting was carried out as in Ravi Ram and Wolfner [68]. Anti-CG11864 antibody [81] was used at a concentration of 1:500 in 5% non-fat dry milk blocking solution, anti-Acp36DE [205] and anti-alpha-Tubulin (Sigma T5168, clone B-5-1-2) at 1:10,000, anti-ovulin [217] at 1:1,000, and anti-Acp62F [76] at 1:2,000. We developed ovulin and CG11864 Western blots with Amersham ECL Prime substrate (GE Healthcare) and all other Westerns with Amersham ECL and exposed the blots to X-ray film. Exceptions are the blots in Figures 3.2A and B and Figure 3.6A which were developed with Amersham ECL Plus and imaged using a Typhoon 8600 imager (GE Healthcare).

### ***Identification of the blood Transposable Element in DGRP-517***

Dr. Margarida Cardoso-Moreira (Cornell University) analyzed publicly available 454 long reads mapping to the CG11864 locus in line 517 from the DGRP sequencing project [219] and identified reads that partially map to CG11864 and partially to a *blood* transposable element. We narrowed the insertion point to between 15,613,145 and 15,613,148 on chromosome 2L, within the first exon of CG11864. The exact insertion point is ambiguous due to an AT repeat at that location.

### ***Ejaculate collections***

We exposed *UAS-dTrpAI/+; fru-GAL4(B)/+* males to 29°C for approximately 5 minutes. After killing males by piercing their thorax, we placed males under a dissecting scope near a

drop of either Buffer A or B (see below). We then collected the ejaculate from the male's lower abdomen using forceps and a small amount of buffer. The ejaculate was generally present as a semi-solid mass containing sperm and seminal fluid, though some fluid may have been lost during collection. The ejaculate was then placed into a tube containing 5 $\mu$ L buffer. Three male ejaculates were pooled per sample. We gently ground samples with a pestle and left them to incubate at room temperature (times indicated on Western blots) before adding SDS sample buffer for Western blot analysis. The following buffers were used:

Buffer A = 50mM Tris-HCl, pH 7.5, 1mM MgCl<sub>2</sub>, 1mM CaCl<sub>2</sub> [234].

Buffer B = 20mM Tris-HCl, pH 7.5, 150mM NaCl, 2.5mM CaCl<sub>2</sub>, 2 $\mu$ M ZnCl<sub>2</sub> [235].

These buffers were chosen because of their previous usage in *in vitro* cleavage assays with astacin-family metalloproteases.

### ***Ovulation Assay***

We mated inbred *w<sup>1118</sup>* females to either CG11864 null or control males and placed them on ice at 3 hours ASM. We placed females in a cold drop of PBS under a dissecting microscope and gently tore the ventral abdomen with sharp forceps to reveal the ovaries, oviduct, and uterus. We recorded the number of eggs in the lateral or common oviducts (0, 1, or 2) [20]. We analyzed the data using a standard T-test in R.

### ***Egg Count Assays***

We performed all assays by mating females to either CG11864 null males or their controls. All experiments were analyzed in R using a standard T-test. There was no significant effect of experimental block, so data collected over several blocks were combined.

10 day egg count assay: These were carried out as in [105]. Briefly, we used 3-6 day old virgin females that had been fed standard yeast/glucose media supplemented with live yeast and allowed them to lay eggs on standard yeast/glucose media after mating. Females laid eggs for 24 hours ASM before being transferred to a new vial and this was repeated for an additional 9 days. We counted the number of eggs laid for each female on each day of the experiment. Egg counts from the first 24 hours were used for the 24 hour egg count assay.

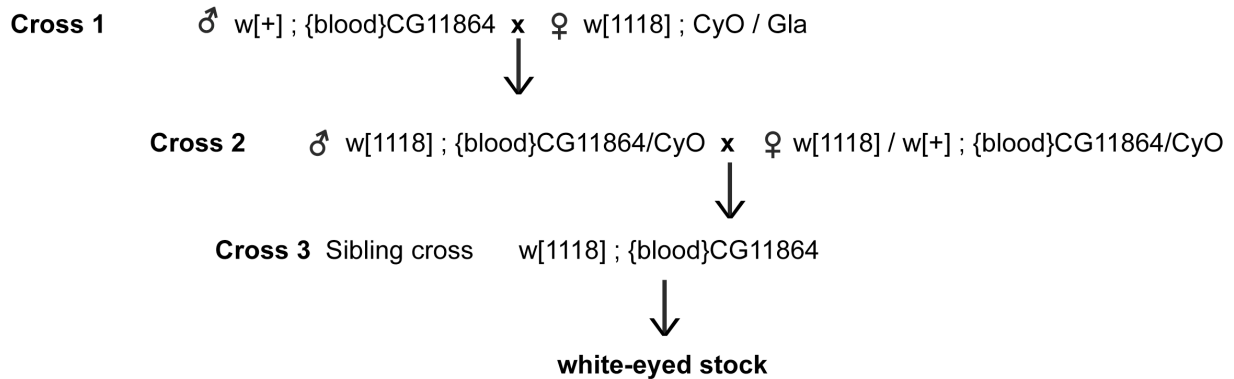
5 day egg count (with 2 day old females): We used 48 hour post-eclosion virgin females fed standard yeast/glucose media (no added yeast). The assay was similar to the 10 day assay, except that eggs were counted only for 5 days.

3 day egg count (with 10-12 day old females): We used 10-12 day old virgin females fed standard yeast/glucose media (no added yeast). The assay was similar to the 10 day assay, except that eggs were counted only for 3 days.

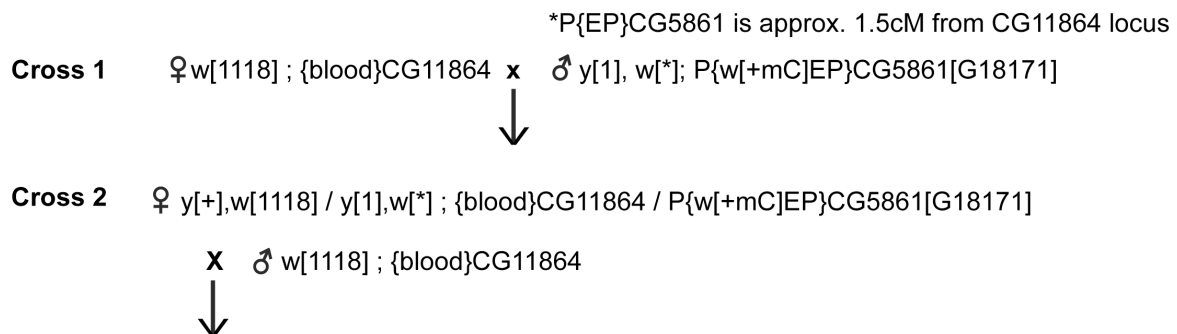
## SUPPLEMENTARY METHODS

### PART 1

1. Remove endogenous w allele from DGRP-517



2. Link null allele to visible marker



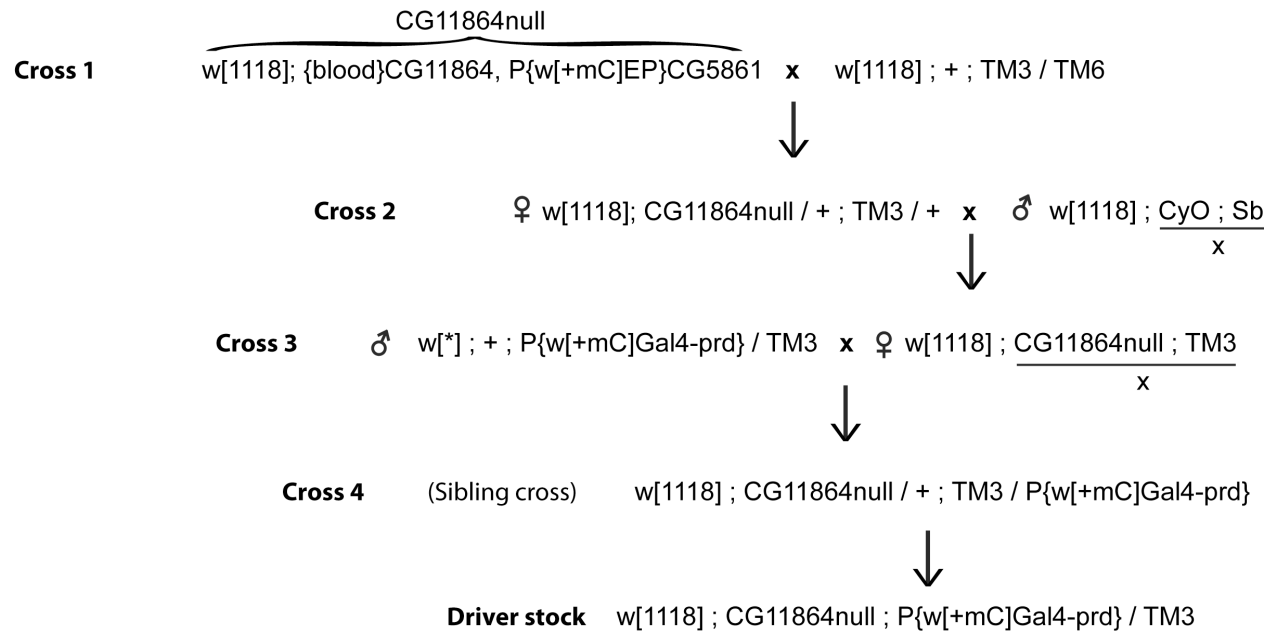
- A. Mate single females to  $w[1118] ; \{blood\}CG11864$
- B. Allow to lay eggs for 24 hours
- C. Analyze for loss of WT CG11864 locus by PCR
  - Extract DNA from single females.
  - PCR with CG11864 primers:  
Forward primer: CTC TTT CTC TAT AGG CAT GTC G  
Reverse primer: GAA GGA GCT CGT GGA TAA TG
  - Look for failed PCR reaction (indicates homozygous for CG11864 null allele). Seminaise PCR used as a control for DNA quality.
  - Identified 1 recombinant in 184 tested.
- D. Make stock homozygous on  $2L\{blood\}CG11864, P\{w[+mC]EP\}CG5861[G18171]$

3. Backcross homozygous stock to  $w[1118]$  for 10 generations (Check at 5 and 10 gen.).
4. In parallel to 3, backcross P{EP}CG5861 stock to  $w[1118]$  for 10 generations.

## SUPPLEMENTARY METHODS

### PART 2

*paired*-Gal4 driver line in CG11864 null background



## CHAPTER 4

### ANALYSIS OF OVULIN PROCESSING RATE VARIATION IN INBRED LINES OF *DROSOPHILA MELANOGATER*<sup>4</sup>

#### INTRODUCTION

The *Drosophila* Genetic Reference Panel (DGRP) is a population of inbred lines with a high degree of natural genetic variation [219]. This panel provides the opportunity to take advantage of this natural variation to study the genetic basis of quantitative traits or to identify new, naturally occurring mutations in pathways of interest. A living library of naturally occurring mutations is especially valuable for traits or pathways that are difficult to dissect using standard genetic screens for any of several reasons. For example, sufficient mutations may not be recovered with standard methods; the mutant phenotype may be too subtle to be identified in a single mutant individual; or, the phenotype may be expressed only under certain conditions (e.g., after mating) that are prohibitive for use in a large-scale screen.

The study of proteolytic processing of seminal fluid proteins is an example of a phenomenon that could benefit from the use of the DGRP. Only two proteases are known to affect the processing of any *Drosophila* seminal fluid proteins: the astacin metalloprotease CG11864 [81] and the trypsin-like serine protease semine [63] are required for post-mating processing of two seminal fluid proteins, ovulin and Acp36DE. Using genetic variation in the rate of processing of these substrate proteins may lead to the identification of other proteins

---

<sup>4</sup> The ovulin cleavage sites used in this chapter were mapped by Dr. Laura Sirot (College of Wooster) with the aid of the Cornell University Proteomics and Mass Spectrometry core facility. Dr. L. Sirot also provided the Western blot image that is presented in Figure 4.1.



involved in this proteolytic pathway, or could uncover informative mutations in pathway components already identified. Here, I have tested for the presence of variation in the processing of the *Drosophila melanogaster* seminal protein ovulin and tested several methods for measuring this variation.

Ovulin is synthesized in the male accessory glands and transferred to females during mating [217]. Ovulin is secreted into the lumen of the accessory glands as a 244-amino acid protein after removal of an N-terminal 20-amino acid predicted secretion-signal peptide. Ovulin exists in at least three glycoforms that differ in their extent of N-linked glycosylation [106]. In the female, ovulin localizes to the oviduct and the base of the ovary, where it functions in release of oocytes from the ovary during the first 24 hours after mating, resulting in a modest increase in eggs laid during this timeframe [20, 105], though the “extra” eggs laid due to ovulin receipt are less likely to be fertilized [236]. Thus, ovulin may function to remove accumulated old oocytes present in the ovaries of virgin females [236]. Ovulin also enters the female’s circulation during the first 10 minutes after the start of mating (ASM) by crossing the posterior vaginal wall [76]; from the circulation, it has access to neural and endocrine tissues from where it could also affect ovulation.

Starting at approximately 10 minutes ASM, ovulin is sequentially cleaved, in three steps, from its amino terminus [106]. This cleavage requires contributions from both the male and female [63, 81, 106] [This thesis, chapter 3]. Ectopic expression of full-length ovulin or either of two of its cleavage products (those produced by the third cleavage step) can each independently stimulate ovulation in virgin females [108]. In an ectopic expression experiment, these two ovulin cleavage products were slightly more potent than full-length ovulin at stimulating virgin ovulation [108], suggesting that the cleavage of ovulin may release additional bioactive products

that increase the level of post-mating ovulation. However, this remains speculative.

As a first step to determining whether ovulin processing rate is sufficiently variable among the DGRP lines to allow for a genetic screen using this trait, I attempted to measure the rate of ovulin processing using three methods: semi-quantitative Western blotting, non-quantitative Western blotting, and quantitative mass spectrometry. Here, I present preliminary data from these three methods and discuss the advantages and disadvantages of each method, along with providing suggestions for future research.

## **METHODS**

### ***Flies***

Fly stocks used in this study include the wildtype Canton-S lab strain and a subset of the inbred *Drosophila* Genetic Reference Panel lines (DGRP). DGRP lines were kindly provided by T. Mackay. All stocks were reared on standard yeast-glucose media at  $23\pm 1^{\circ}\text{C}$  on an approximately 12h/12h light/dark cycle.

I observed single-pair matings between 3-5 day old virgin males and females of the same inbred line in glass vials. Each vial contained a moistened square of filter paper for hydration. I discarded matings that lasted less than 10 minutes. At the appropriate time ASM, single females were flash-frozen in liquid nitrogen in 0.5-mL plastic tubes and kept at  $-20^{\circ}\text{C}$  until dissection.

### ***Sample preparation***

I dissected female lower reproductive tracts (uterus and sperm storage organs, RT) under a dissecting microscope. For Western blot samples, I used 1x PBS as the dissection and grinding

buffer. I used trypsin digestion buffer for samples submitted for mass spectrometry. I placed tissues in approximately 10 $\mu$ L buffer in a 0.5-mL plastic tube and briefly ground them with a pestle. For Western blot samples, I then added 10 $\mu$ L of 2x SDS sample buffer containing 10% 2-mercaptoethanol to the samples, which were then boiled for 5 minutes.

### ***Quantitative Western blots***

Female RT samples were prepared as above from females frozen at 30' ASM. All females were mated to males from their same inbred line. Two females were pooled for each sample. I separated proteins on Tris-glycine gels containing 5-20% acrylamide (gradient gels were prepared individually by hand). Western blotting was carried out as in Ravi Ram and Wolfner [68]. I used anti-ovulin antibody [217] at a concentration of 1:1000 in 5% non-fat dry milk blocking solution. I developed ovulin Western blots for quantitative analysis with the fluorescent/chemiluminescent reagent ECL-Plus (Amersham, GE Healthcare) and imaged them using a Typhoon 8600 imager (GE Healthcare). Measurements of band intensities were carried out using the program ImageQuant (GE Healthcare).

I used the following statistic to represent the extent of ovulin processing for these lines: the difference in intensity of the ~22kDa band (corresponding to CP3) and the ~30kDa band (corresponding to CP2) was divided by the total intensity of all ovulin bands in the sample. This statistic (Q, for quantification) most accurately reflects the rate of the final cleavage event (CP2 to CP3) (see [108] for terminology).

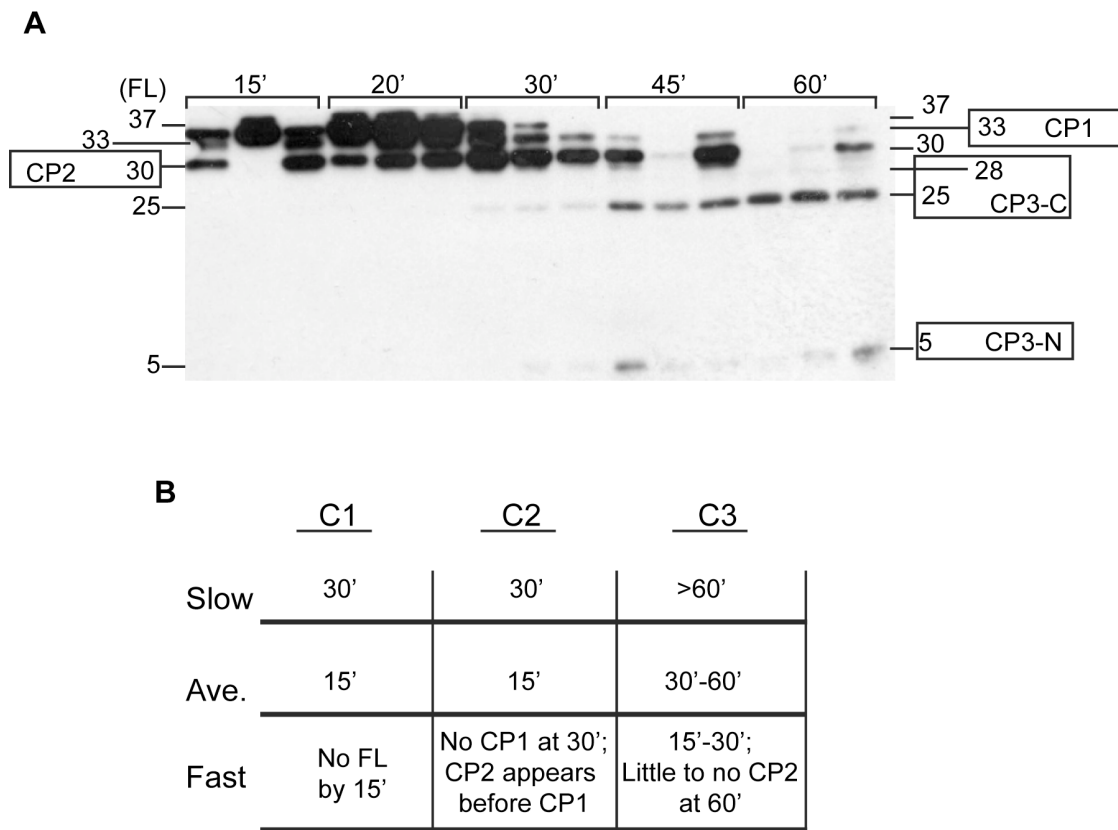
### ***Non-quantitative Western blotting***

Female RT samples were prepared as above from at least two females per line frozen at

each of three timepoints: 15', 30', and 1 hour ASM. All females were mated to males from their own inbred line. Each sample contained protein from a single female RT. I separated proteins from the RT extracts using SDS-PAGE with 12% polyacrylamide Tris-glycine gels for approximately 3 hours. Western blotting was carried out as in Ravi Ram and Wolfner [68]. After transfer to PVDF membrane, the membranes were cut to separate the ovulin- and Tubulin-containing portions of the membrane and allow for simultaneous incubation with anti-ovulin and anti-Tubulin antibodies without introducing background noise due to an interaction between the antibodies. I used anti-ovulin antibody [217] at a concentration of 1:1000 in 5% non-fat dry milk blocking solution, and anti-alpha-Tubulin (Sigma T5168, clone B-5-1-2) at 1:10,000. With Westerns for qualitative analysis, I used the more sensitive chemiluminescent reagent ECL prime (Amersham, GE Healthcare), in order to detect signal from single females, and exposed the blots to X-ray film.

### ***Ovulin processing rate categories for qualitative analysis***

I examined ovulin from at least 2 females per line at each of three timepoints (15', 30', and 1 hour ASM) to determine the extent of ovulin processing. Each Western blot lane contained protein from one mated female RT. Figure 4.1A presents a Western blot showing the progression of ovulin processing in single, mated CS females along with the names and corresponding sizes of the different cleavage products indicated.



**Figure 4.1: Ovulin cleavage products and rate categories. A)** Western blot, probed with anti-ovulin antibody (image credit: L.Sirot). Each lane contains a single female reproductive tract (RT) at the time indicated after the start of mating (ASM). All samples are of CS females mated to CS males. Numbers next to the blot indicate approximate size in kDa. FL=full-length (36-41kDa); CP1 = first cleavage product (33-36kDa); CP2 = second cleavage product (30-33kDa); CP3-C = third cleavage product (25-28kDa). The third cleavage event also produces a 5kDa product (CP3-N) not used in this study. Cleavage product sizes were determined by Park and Wolfner, 1995 [106]. **B)** Rate categories used in this study. If only time is indicated (time ASM) this is the first timepoint at which the cleavage product is detected in at least 2 samples. Timepoints used for this study: 15', 30', and 60' ASM.

I rated each cleavage product (CP1, CP2, and CP3-C) based on the timepoint at which it first appeared and/or when it last appeared. For the purposes of this study, “average”, “fast”, or “slow” cleavage event 1 (C1), cleavage event 2 (C2) or cleavage event 3 (C3) refers to the inferred speed of the first, second and third cleavage events, respectively, which produce the three visible products (CP1, CP2, and CP3-C). A summary of the criteria I used is shown in Figure 4.1B. For a line to be categorized as average, both the first and second cleavage products must be visible at 15’ ASM and the third cleavage product visible either at 30 or 60 minutes ASM. The third cleavage product is not strongly recognized by the ovulin antibody, making it difficult to detect before 60’ ASM.

Fast C1 was defined as having no or very little visible full-length protein at 15’ ASM. Fast C2 was defined as there being no visible CP1 by 30’ ASM, or having visible CP2 at 15’ or 30’ ASM when no or very little CP1 is ever present. The second scenario would occur when C1 is slow, but C2 is fast; slow C1 is determined by presence of full-length ovulin and absence of CP1 at 15’ ASM.

CP3 in the fast category is visible at 15 or 30’ ASM and/or is more prominent than the second cleavage product at 60’ ASM. The third cleavage event was never observed to be “slow”. This may be partly due to the fact that our antibody does not efficiently detect CP3 [106]. Thus, a low intensity CP3 band at 60’ ASM does not necessarily indicate slower processing of CP2 to CP3. For this reason, I would only categorize C3 as slow if CP3 was never present by 60’ ASM and CP2 was still clearly present at 60’ ASM. No line met these criteria in this screen. I have presented all of the Western blot data used in this analysis at the end of this chapter (Figure 4.9).

### ***Quantitative mass spectrometry: Peptide selection and verification***

Five female RTs were prepared as above at the times ASM given for each experiment. Samples were digested with chymotrypsin and prepared for mass spectrometry by the Cornell Proteomics and Mass Spectrometry core facility at Cornell University. I worked with the core facility to develop an assay using selected reaction monitoring (SRM) mass spectrometry [237, 238]. SRM mass spectrometry uses a triple quadrupole mass spectrometry instrument with the first and third quadrupole acting as mass filters (the second is a collision cell) to monitor the abundance of a pre-selected peptide precursor ion and ion fragment. The specific precursor ion and fragment pair is known as a transition. By monitoring one or several transitions, using the known retention time in the instrument of the desired peptide, one can get a reliable quantification of that peptide's abundance in a complex sample, even if the peptide is of relatively low abundance in the sample.

The first mass filter selects for the exact mass of the pre-selected peptide ion, which must be determined empirically with standard shotgun mass spectrometry methods. The second mass filter selects a particular ion fragment produced after fragmentation of the precursor ion in the second quadrupole. Over the course of the experiment, the instrument alternately scans through several transitions (precursor/fragment pairs) and produces a chromatograph of the instrument retention time and signal intensity for each transition. The mass filters allow the complexity of the sample to be reduced and thus produces reliable measurements of the desired peptides. When multiple peptides are monitored in a single experiment, the technique is often called “multiple reaction monitoring”, or MRM. Transitions monitored in each experiment are shown in the supplementary materials.

While ovulin itself is fairly abundant in the mated female [57], its individual cleavage products are not easily detected in the extremely complex protein mixture that is the mated

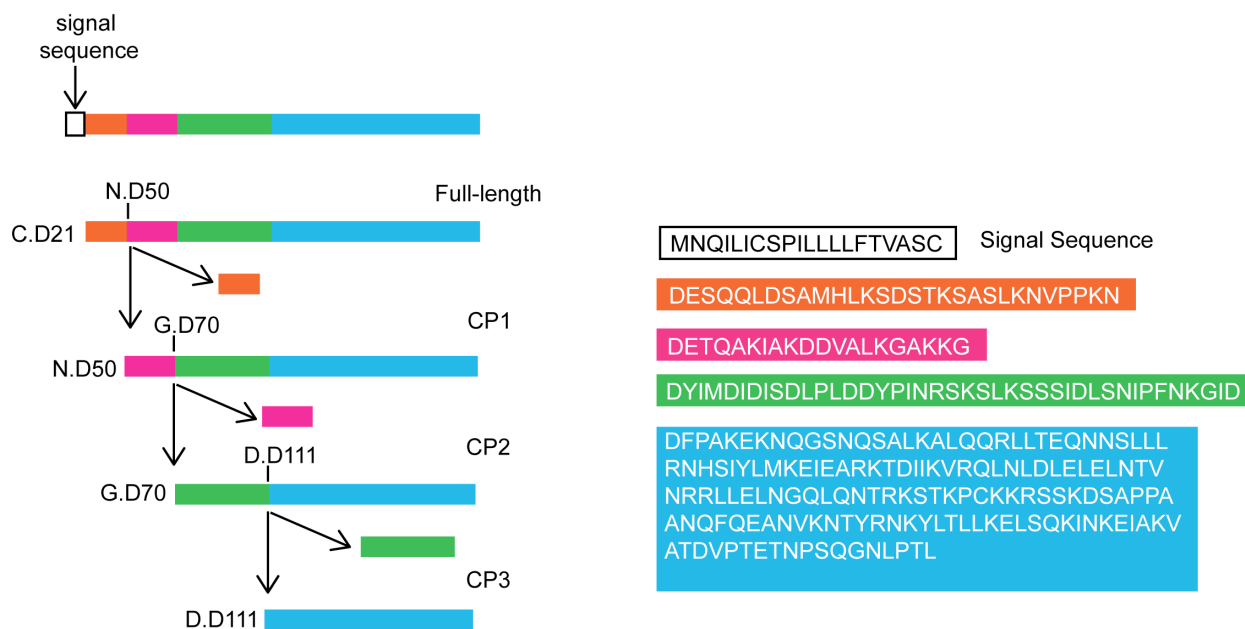
female RT. This necessitated the use of SRM instead of standard shotgun mass spectrometry to quantify the cleavage products. A former postdoc in the Wolfner lab, Dr. Laura Sirot mapped the cleavage sites of ovulin (unpublished data; Figure 4.2). Using these data, obtained from shotgun mass spectrometry experiments on isolated cleavage product bands from 1D SDS-PAGE, two peptides were chosen to monitor the rate of the third ovulin cleavage event (peptides specific to the first and second cleavage products were not found to be suitable for this experiment). The two peptides were a peptide specific to the third cleavage product (“cleaved CP3”: DFPAKEKNQGSNQSAL), and a peptide that spans the third cleavage product, and is thus absent after cleavage (“uncleaved CP3”: NKGLDDFPAKEKNQGSNQSAL).

The pilot experiment to validate the selected peptides was done with a 2-hour gradient on an LTQ Orbitrap Velos. Lower RTs from CS females mated at 1 hour ASM were prepared and digested with chymotrypsin. This sample was run through the instrument in data-dependent acquisition mode, meaning that a full mass spectrum was recorded, before SRM. The mass spectra were needed to confirm that the peaks analyzed with SRM did in fact originate from the desired peptide. For the uncleaved CP3 peptide, two transitions were monitored in this test run. One transition was monitored for the cleaved CP3 peptide. Both peptides were deemed suitable for analysis by SRM.

A second pilot experiment was performed to confirm that the cleaved CP3 peptide was detected in samples where it should be abundant (e.g., in females at 1 hour ASM), but not in samples where the third cleavage event should not have occurred (e.g., in females at early times ASM, such as 10', or in females mated to CG11864 null males [Chapter 3]). I submitted three samples for analysis: CS females mated to CS males at 10' ASM, CS females mated to CS males at 1 hour ASM, and CS females mated to DGRP-517 males (CG11864 null) at 1 hour ASM. A



third peptide was monitored in this experiment: a C-terminal control peptide that should not be affected by ovulin cleavage (LMKEIEARKTDIIKVRQL). Four transitions were monitored for the control peptide, two for the uncleaved CP3 peptide, and four for the cleaved CP3 peptide. The average of all transition peaks was used as the raw measurement for each peptide's abundance.



**Figure 4.2: Ovulin cleavage sites determined by Laura Sirot.** Ovulin cleavage sites and sequences of ovulin cleavage products as determined by mass spec analysis performed by Dr. Laura Sirot (unpublished work). Ovulin is cleaved from the N-terminus in three sequential steps. Ovulin is shown at the top left with its secretion-signal sequence intact (pre-secretion ovulin). After removal of the signal peptide, the N-terminal residue is D21 (numbering based pre-secretion ovulin). This form is referred to as “full-length”; this is the form of ovulin transferred to females during mating. Arrows below each protein schematic indicate point to C-terminal (left) and N-terminal (right) cleavage products produced after each cleavage step. Color-coded sequences corresponding to each cleavage product are shown on the right.

I applied two methods of normalization to analyze the data. The first was to normalize the peptides from the CS 1h and 517 1h samples to the corresponding peptide in the CS 10' sample, by setting the CS 10' samples all to 1. The second method normalized the cleaved and uncleaved CP3 peptides within each sample to the control peptide (setting the control peptide abundance to 1).

### ***Pilot experiment with two DGRP lines***

I chose to analyze DGRP lines 786 and 852, based on the results in my pilot quantitative Western blot study. Females were mated singly to males of the same DGRP inbred line and frozen at 30' ASM (the same time used in the quantitative Western screen). The samples were each run on two separate instruments. The first was the same instrument (Orbitrap) used in the previous pilot experiment, again in data-dependent acquisition mode. The second instrument (4000 Q Trap triple-quadrupole linear ion trap) was newly acquired by the mass spectrometry facility and is more sensitive than the Orbitrap. This was used in SRM-triggered MS/MS mode to obtain quantitative data. In the Orbitrap experiment, 2 transitions were monitored for the control peptide, and one each for the cleaved and uncleaved CP3 peptides. In the Q Trap experiment, 7 transitions were monitored for the control peptide, 3 for the uncleaved CP3 peptide, and 4 for the cleaved CP3 peptide

## **RESULTS AND DISCUSSION**

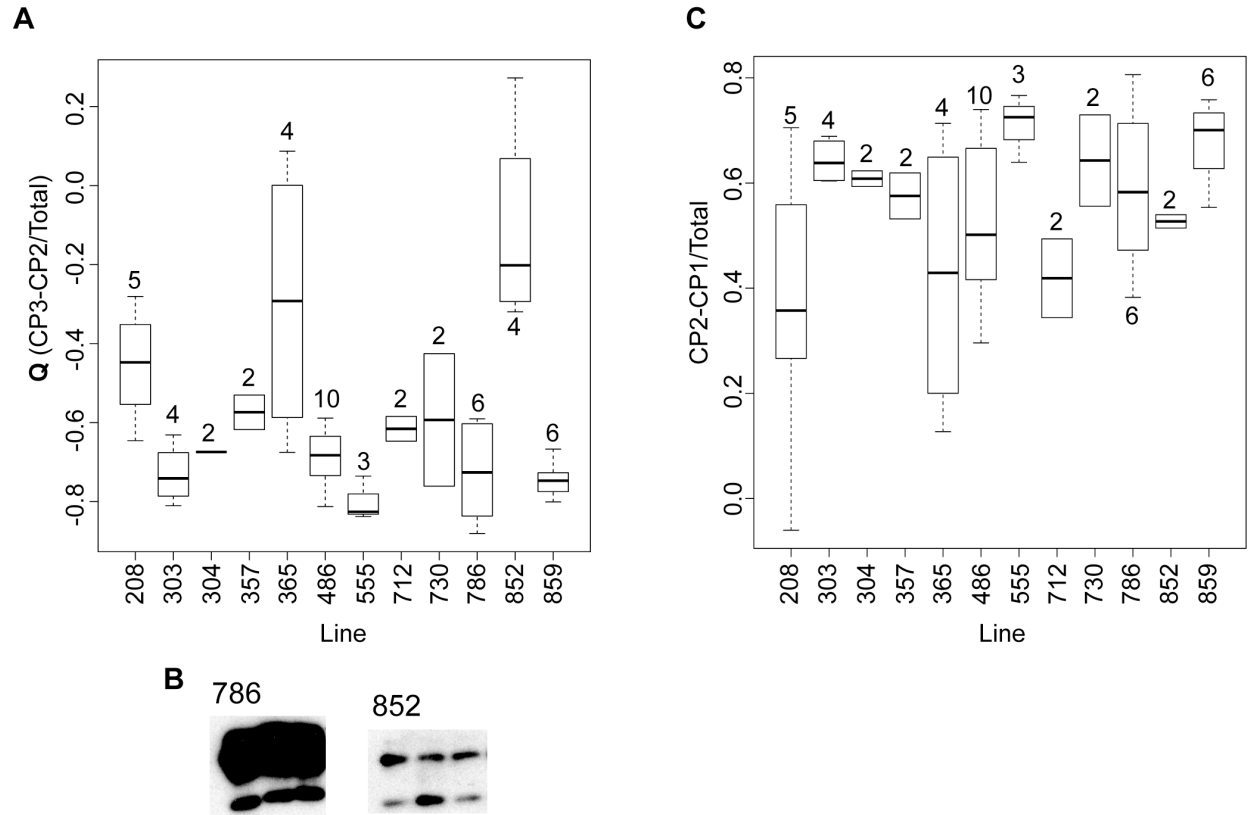
### ***Ovulin processing at 30' ASM varies quantitatively between DGRP lines***

In a standard laboratory *Drosophila* strain (Canton-S, or CS), ovulin is processed to

completion by about 1 hour after the start of mating (ASM). The first and second cleavage products are visible on a Western blot by 15' ASM, the third by about 30' ASM (Figure 4.1A). By 1 hour ASM, only the third cleavage product(s) are usually visible, though a small amount of earlier cleavage products may remain. While these observations are generally true in the CS strain, considerable variation does exist between individual mated females for the rate of ovulin processing. I therefore chose to examine ovulin's processing rate in a subset of the DGRP inbred lines, by mating females to males of their own line, to determine if there is genetic variation for the rate of ovulin processing.

For this initial study, I only used females dissected at 30' ASM, as this was the timepoint with the most variation in ovulin processing extent (Figure 4.1A). I calculated the Q statistic (see Methods) as a proxy for the extent of ovulin processing in each sample. Q reflects the rate of the third, and final, cleavage event by measuring the difference in intensity between CP3-C and CP2 (See Figure 4.1A for cleavage product sizes; Supplementary Figure 4.2 shows a schematic of ovulin processing).

There was significant variation in the Q statistic among the 12 lines that yielded acceptable Western blots for at least 2 independent biological replicates ( $F = 7.43$ ,  $p < 0.0001$ ) (Figure 4.3A). Two lines, 786 and 852, were selected for analysis by quantitative mass spectrometry (see section on SRM mass spectrometry). Representative ovulin Western blots of females from these two lines at 30' ASM are shown in Figure 4.3B. A second statistic reflecting the difference in intensity between CP2 and CP1 was not significantly different between lines ( $F = 1.74$ ,  $p = 0.103$ ) (Figure 4.3C).



**Figure 4.3: Quantification of ovulin processing rate as measured using Western blots for 12 inbred lines.** **A)** Ovulin processing rate for 12 DGRP inbred lines. Processing rate is defined as the difference in intensity of CP3 and CP2 normalized to the total sample intensity. Samples were prepared from 2 female RTs at 30 minutes ASM. ANOVA:  $F = 7.43$ ,  $p < 0.0001$ . **B)** Representative Western blots from lines 786 and 852 (used later in mass spec pilot study). Each blot shows 3 identically prepared samples at 30' ASM. Line 786 had a slower measured processing rate than 852, as determined by a lower Q value. **C)** Processing rate for the same samples as in **A**, using a different statistic ( $CP2 \text{ intensity} - CP1 \text{ intensity} / \text{Total}$ ). ANOVA:  $F = 1.74$ ,  $p = 0.103$ . Numbers above or below boxplots indicate sample size.

These data suggest that the rate of ovulin processing has a genetic basis. Factors such as proteases, proteolysis co-factors, post-translational modification of ovulin, and even pH of the female RT may explain the variation seen among these lines. Further work is required to determine the relative contributions made to ovulin processing rate by males and females.

Measuring rates from matings between a standard lab strain male or female mated to DGRP flies

of the opposite sex would help establish the relative contributions of each sex.

This Western blot-based approach allows for quantitative measurements of ovulin processing products to determine a single measurement of ovulin processing rate. It is even possible to obtain measurements for each cleavage event separately, and at multiple timepoints ASM, to create a more detailed understanding of ovulin processing rate. This approach may also be useful for other seminal fluid proteins that are proteolytically processed after mating, such as Acp36DE [7, 205].

One consideration for future use of this method is the normalization approach. Here, I have used the total ovulin signal intensity to normalize the CP3-CP2 difference. This may not be the optimal method, as the ovulin antibody is known to have different sensitivities for each cleavage product, and this will bias the results. More fully processed samples would be expected to have lower signal intensities overall, since CP3 is the least antigenic cleavage product. Thus, these samples would have an artificially high Q value. Ideally, a suitable “house-keeping” protein would be identified that does not differ significantly between the DGRP lines.

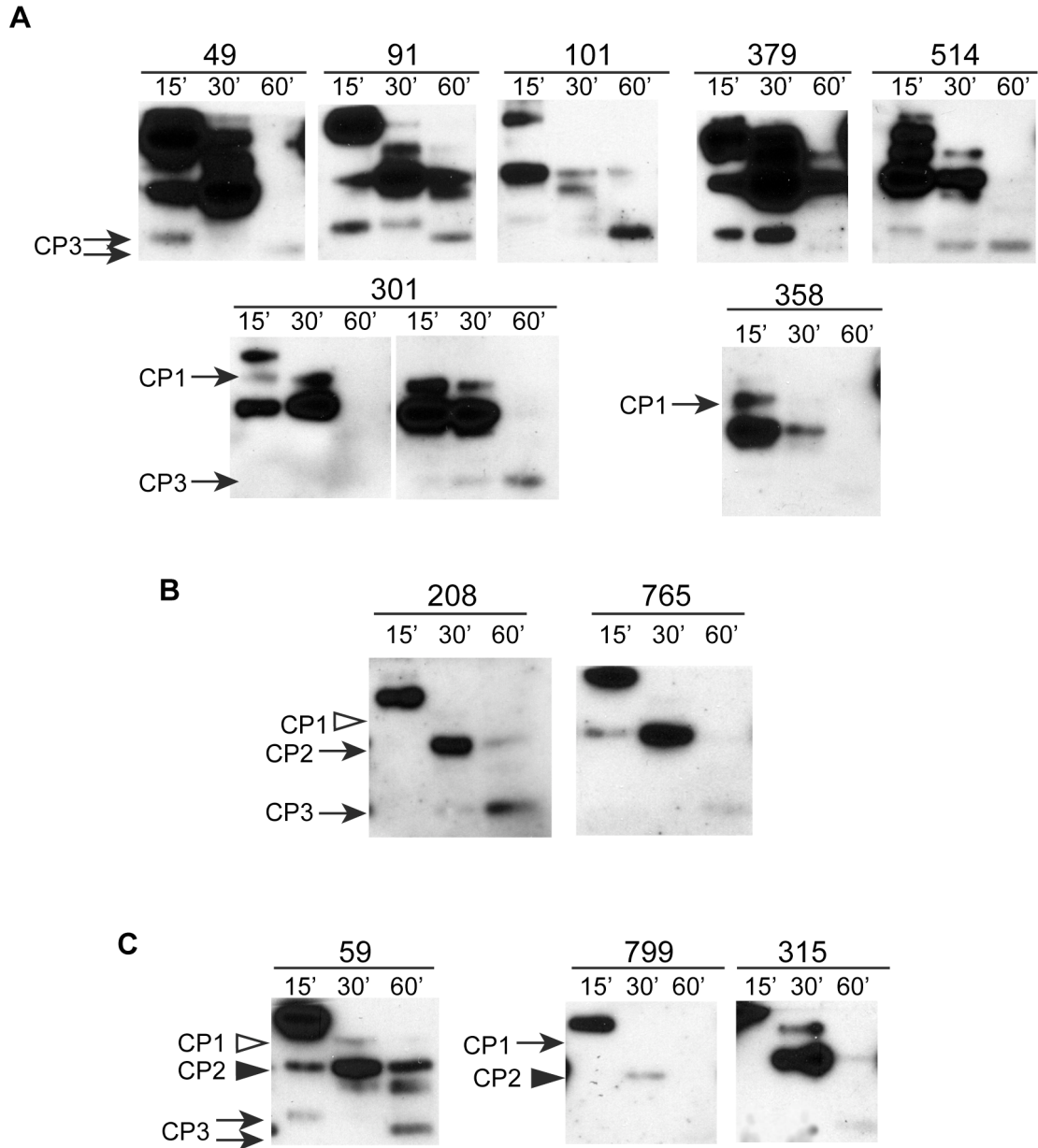
### ***A qualitative approach finds the rate of individual ovulin cleavage products varies among 25 DGRP lines***

The quantitative Western blot-based approach described above revealed that there was likely genetic variation for ovulin processing rate. However, a disadvantage of this approach is that only a small proportion of Western blots are of high enough resolution to use for quantification. Low signal-to-noise ratios and splotches obscuring the ovulin bands precluded the majority of samples from being used in this analysis. A non-quantitative, categorical approach could be used with lower-quality Western blots, as only the presence/absence of bands or gross

relative intensities of processing products need be noted. I therefore developed a categorical system for qualitatively describing the ovulin processing rate for the three cleavage events in each line (see Methods).

For each cleavage product, most lines fell into the “average” rate category (Table 4.1). Figure 4.4 shows representative blots from each line that had either fast or slow processing for at least one cleavage event (the complete set of Western blots used is shown in supplementary Figure 4.9). A cleavage product was categorized as either fast or slow if at least 2 independent samples showed the same pattern of either fast or slow. However, C1 in line 301 was an exception, as it only showed a fast C1 pattern in 1 independent sample, but not the other. More samples are needed from this line to confirm fast C1.

C3 was the most prevalent “fast” event (Table 4.1). Six lines had fast C3 and average rates of other cleavage events (Table 4.1 and Figure 4.4A). Lines with fast C3 tended to have visible CP3 at 15’ ASM, but only the 28kDa glycoform was detectable. At 30 and 60’ ASM, the 25kDa glycoform was the predominate form of CP3. This is clearly visible, for example, in line 91 (Figure 4.4A), in which CP3 is visible as the 28kDa glycoform at 15’ ASM, but at the smaller 25kDa form at later times ASM. These data indicate that different glycoforms of ovulin may be processed at different rates (previous work indicated that the glycoforms were processed at similar rates, though this was only demonstrated in one lab strain and not in individual females [106]).



**Figure 4.4: DGRP lines with fast or slow ovulin processing.** Western blots, probed with anti-ovulin antibodies. Each lane contains protein from a single mated female at the indicated time ASM. CP3 here indicates CP3-C [106]. **A)** DGRP lines with fast processing. A single representative ovulin Western blot is shown for each line (2 blots shown for line 301). Arrows indicate relevant cleavage product(s). **B)** DGRP lines with slow first cleavage event, but fast second and third cleavage events. Arrows indicates cleavage products with fast processing; white arrowhead indicates cleavage product with slow processing. **C)** DGRP lines with slow C1 and an ambiguous rate of C2. White arrowheads=slow; black arrowheads: rate could not be determined (see text); arrows: fast.

**Table 4.1: Ovulin processing rate variation in 25 inbred lines.** DGRP Lines with at least 2 usable samples were analyzed for rate of processing of the three ovulin cleavage events. Fast or slow refers to time at which that cleavage product first appeared (fast=earlier than average; slow=later than average). CP3 could not be detected for lines 161 and 799. Average samples are denoted by dashed lines. NA=could not be determined.

<b>DGRP Line</b>	<b>Sample N</b>	<b>C1</b>	<b>C2</b>	<b>C3</b>
28	2	--	--	--
45	1	NA	NA	NA
49	2	--	--	<b>fast</b>
57	4	--	--	--
59	2	<b>slow</b>	<b>fast?</b>	<b>fast</b>
91	2	--	--	<b>fast</b>
101	2	--	--	<b>fast</b>
161	2	--	--	<b>fast?</b>
176	2	--	--	--
208	2	<b>slow</b>	<b>fast</b>	<b>fast</b>
272	2	--	--	--
301	2	<b>fast</b>	--	<b>fast</b>
304	3	--	--	--
315	4	<b>slow</b>	<b>fast?</b>	--
319	4	--	--	--
324	3	--	--	--
358	2	<b>fast</b>	--	--
365	2	--	--	--
379	2	--	--	<b>fast</b>
437	2	--	--	--
514	3	--	--	<b>fast</b>
765	4	<b>slow</b>	<b>fast</b>	<b>fast</b>
774	2	--	--	--
786	2	--	--	--
799	2	<b>slow</b>	<b>?</b>	<b>fast?</b>

Two lines (301 and 358) had fast C1 (Table 4.1 and Figure 4.4A). Another two lines (208 and 765) had slow C1 and fast C2 and C3. These lines had no visible CP1 at either 15' or 30' ASM, but CP2 was present at one or both of these timepoints (Figure 4.4B), suggesting that CP1 had already been cleaved to produce CP2. CP3 then appeared rapidly as CP2 was depleted



(Figure 4.4B).

Three additional lines, 59, 799, and 315, also had fast C1, but I could not clearly determine if C2 was also fast (Figure 4.4C). This was either because it was unclear whether CP1 was present at 15' due to strong full-length protein signal (line 59); CP1 and CP2 were difficult to detect at all timepoints (799); or CP1 and CP2 both appeared at 30' ASM but not before (3 out of 4 Line 315 samples). Additional samples and/or timepoints need to be examined for these lines. Line 59 also had fast C3 in addition to slow C1 (Figure 4.4C).

Interestingly, the rate of each cleavage event does not appear to be correlated with the rate of processing of other cleavage events across lines. Line 358 had fast C1 (Figure 4.4A), but this was not accompanied by faster C3. Determination of C2 rate would require checking this line at 10' ASM, since CP2 is already present at 15' ASM. In two lines, 208 and 765, the first cleavage event was slower than average (Figure 4.4C), but C3 then proceeded quickly. These data suggest that in some lines, the first cleavage event is the rate-limiting step, while the third may be the rate-limiting step in other lines, but that processing always proceeds sequentially from C1 to C3. This variation may reflect the differences in ovulin's sequence between lines (which is known for the majority of the DGRP lines used here) or post-translational modifications that may alter its structure and thus the availability of the cleavage sites to proteases (e.g., CG11864). Protease efficiency or expression of co-factors for proteolysis may also vary between these lines.

Here, I have used only 2-4 females from each line. This was sufficient to determine that there was variation between lines, though more variation may exist within lines than what I observed here. With additional samples to account for within-line variation, this qualitative screen of ovulin processing rates would yield a useful dataset for the identification of factors

affecting ovulin processing. However, there may still not be enough power for a whole genome association study, given the potentially large amount of information compressed into the rate categories. Thus, this approach would be best geared toward identification of functional mutations in genes already known to be associated with ovulin processing, such as ovulin itself, CG11864 [81], and seminare [63]. One could also use a candidate gene approach by searching the DGRP sequences for mutations in other Sfp genes that correlate with ovulin processing rate.

For future research, a logical approach might be to screen through the sequences of these genes first, then test for defects in ovulin processing rate only in lines with mutations that suggest such a problem. The disadvantage with this approach is that informative mutations may be missed if they were left out of the final genome assembly (as with line 517, see Chapter 3 of this thesis) or if SNPs were not called correctly. However, this approach would still be less labor intensive, and may thus be worthwhile to pursue.

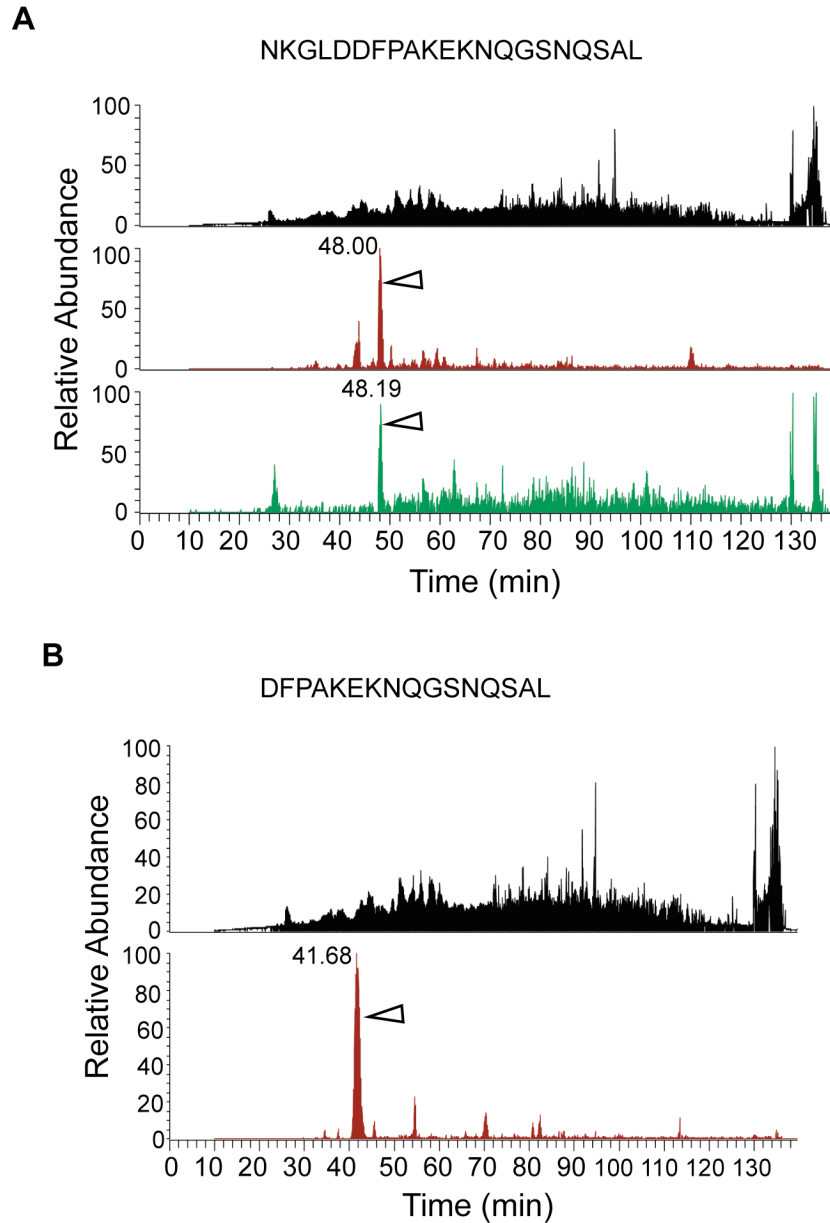
A follow-up question to this screen that should be addressed is the relative contribution that each sex makes to the rate in ovulin processing, as briefly mentioned above. In the present study, I used males and females from within the same inbred line for each mating pair. Using a standard male to mate to females from the DGRP lines, an analogous experiment could determine whether female contributions, such as potential proteases or co-factors, affect the rate of ovulin processing. The inverse approach—standard female mated to DGRP males—would help define the range of male variation for this trait. Varying a single sex at a time may also be a better approach for genomic association studies, as less information will be compressed into the rate measurements/categorizations.

## OVULIN CLEAVAGE PRODUCT QUANTIFICATION WITH SRM MASS SPECTROMETRY

### *Ovulin cleavage products can be detected in complex samples by selected reaction monitoring (SRM) mass spectrometry*

To better quantify the relative amounts of each cleavage product in the mated female RT at a given time ASM, I worked with members of the Proteomics and Mass Spectrometry core facility to develop an assay for measuring ovulin cleavage products (see Methods). This assay is more cost-intensive than the Western blot approaches, but it avoids any possible bias resulting from line-to-line variation in the recognition of ovulin epitopes by the antibody. An initial pilot experiment determined that our selected peptides could be identified in complex samples (i.e., protein extracts from mated female RTs). Figure 4.5 shows the peaks for peptides corresponding to uncleaved CP3 (Figure 4.5A, arrowheads) and cleaved CP3 (Figure 4.5B, arrowhead).

A second pilot study was run to determine whether an increase in ovulin cleavage products could be detected in samples from females dissected at later timepoints ASM relative to those dissected at earlier timepoints. Three samples were submitted for analysis: **1.** CS females mated to CS males at 10' ASM; **2.** CS females mated to CS males at 1 hour ASM; **3.** CS females mated to DGRP-517 males (CG11864 null males [Chapter 3]) at 1 hour ASM. I expected to see the cleaved CP3 peptide as much more abundant in the second sample relative to the other two. Ovulin cleavage is not visible by Western blot until after 10' ASM [106] and females mated to DGRP-517 males should not process any ovulin [Chapter 3]. The full dataset from this experiment is shown in Table 4.2.

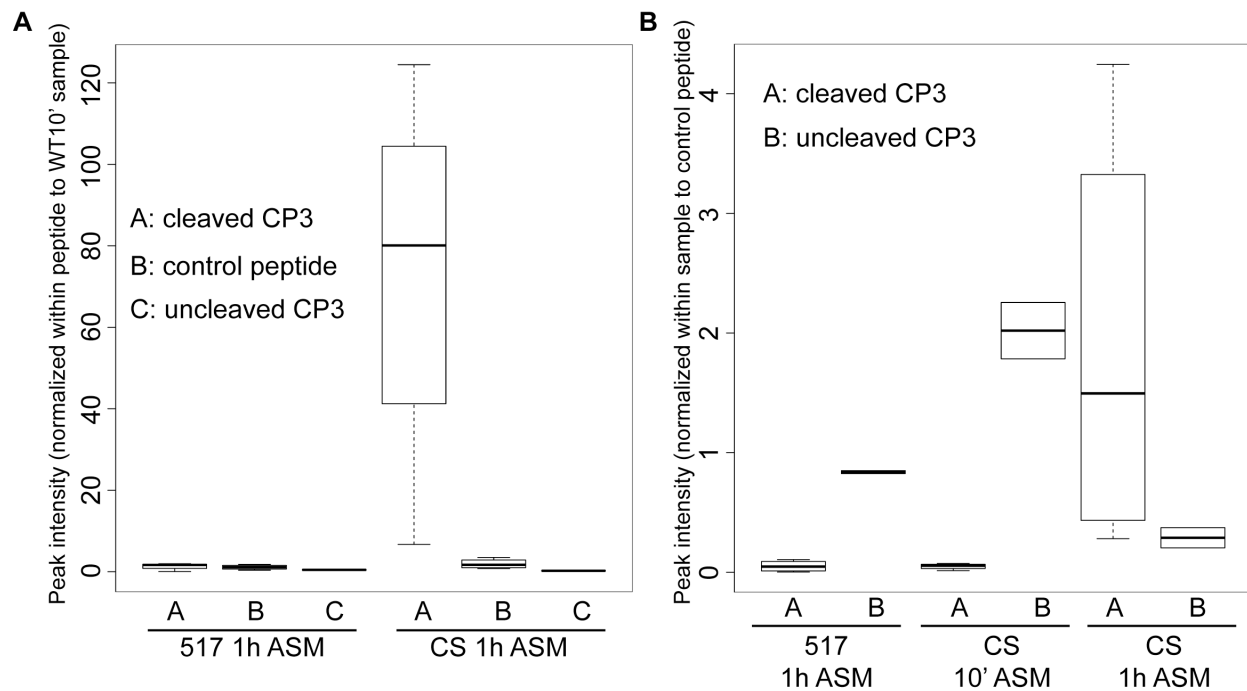


**Figure 4.5: Ovulin CP3 is detected in mated females by SRM mass spec.** Peaks associated with selected ovulin peptide ions. Y-axis: relative peak height; X-axis: retention time on column (in minutes). Black: all peaks for selected peptide; Red: peaks for +2 peptide ion; Green: peaks for +3 peptide ion. **A)** Peptide spanning cleavage products CP2 and CP3 representing amount of ovulin not yet cleaved at third cleavage site. **B)** Peptide specific to CP3 after third cleavage event. Arrowheads indicate best transition ion to use based on their specificity to the desired peptides.

**Table 4.2: Ovulin peptides from females mated to WT (CS) or 517 males.** Ovulin peptides from CS females mated to CS or DGRP-517 males. CS samples are at 10 min. or 1 hour ASM. DGRP-517 samples are from 1 hour ASM. Peptide sequences corresponding to each peptide ID are shown below. **a:** mass-to-charge (m/z) ratios for each precursor/fragment pair (each number represents a m/z ratio). **b:** Charge of peptide ion. **c:** Area of the corresponding peak. **d:** Height of the corresponding peak. **e:** Column retention time of the transition ion. **Control:** LMKEIEARKTDIIKVRQL; **Uncleaved CP3:** NKGLDDFPAKEKNQGSNQSAL; **Cleaved CP3:** DFPAKEKNQGSNQSAL

Sample Name	Peptide ID	<sup>a</sup> Precursor / Fragment	<sup>b</sup> Peptide ion	<sup>c</sup> Area	<sup>d</sup> Height	<sup>e</sup> Retention time
WT10min	Control	728.8/869.5	3+	7185	270.3	49.71
WT10min	Control	728.8/1315.4	3+	1330	45.93	49.52
WT1h	Control	728.8/869.5	3+	9497	320.9	48.33
WT1h	Control	728.8/1315.4	3+	1036	36.40	48.37
5171h	Control	728.8/869.5	3+	8674	252.7	49.14
5171h	Control	728.8/1315.4	3+	667.9	17.86	49.35
WT10min	Control	1092.6/869.5	2+	6003	150.9	49.49
WT10min	Control	1092.6/1315.4	2+	3177	69.12	49.64
WT1h	Control	1092.6/869.5	2+	10410	339.2	48.29
WT1h	Control	1092.6/1315.4	2+	7662	243.4	48.30
5171h	Control	1092.6/869.5	2+	7589	198.6	49.00
5171h	Control	1092.6/1315.4	2+	4864	125.9	49.08
WT10min	Uncleaved	754.4/1010.0	3+	16010	302.3	41.26
WT10min	Uncleaved	754.4/643.3	3+	13660	239.2	41.26
WT1h	Uncleaved	754.4/1010.0	3+	4758	87.61	39.81
WT1h	Uncleaved	754.4/643.3	3+	4339	47.87	40.13
5171h	Uncleaved	754.4/1010.0	3+	11580	126.4	41.21
5171h	Uncleaved	754.4/643.3	3+	1396	122.7	41.06
WT10min	Cleaved	867.4/810.0	2+	4747	6.69	37.61
WT10min	Cleaved	867.4/736.4	2+	7285	8.016	37.65
WT10min	Cleaved	867.4/1175.6	2+	453.6	1.82	37.38
WT10min	Cleaved	867.4/1058.3	2+	254.4	9.824	37.25
WT1h	Cleaved	867.4/810.0	2+	66900	564.8	36.57
WT1h	Cleaved	867.4/736.4	2+	1.132E5	997.6	36.54
WT1h	Cleaved	867.4/1175.6	2+	14890	137.9	36.52
WT1h	Cleaved	867.4/1058.3	2+	7800	66.00	36.56
5171h	Cleaved	867.4/810.0	2+	156.7	11.03	35.66
5171h	Cleaved	867.4/736.4	2+	289.8	15.78	35.15
5171h	Cleaved	867.4/1175.6	2+	27.5	3.122	35.38
5171h	Cleaved	867.4/1058.3	2+	3.905	0.3156	37.09

After normalization of the abundance measurements of the two 1 hour ASM samples to the CS 10' ASM sample, I determined that the cleaved CP3 peptide was highly abundant in the CS 1 hour sample relative to the other two samples, as expected, while the control and uncleaved CP3 peptides did not differ between the three samples (Figure 4.6A). A simple ANOVA found a statistically significant difference in the level of the cleaved CP3 peptide between the 517 1 hour and CS 1 hour sample ( $F = 6.40$ ,  $p = 0.027$ ). There was no statistical difference between the samples for the other two peptides.



**Figure 4.6: Ovulin CP3 is detected at highest levels in WT matings at 1 hour ASM.**

Boxplots show data for each peptide across multiple transition ions. Samples were prepared from CS females mated to the indicated male. **A)** Data normalized within peptide relative to CS 10' ASM. ANOVA (cleaved CP3 peptide only):  $F = 8.54$ ,  $p = 0.027$ . ANOVA (all peptides):  $F = 6.40$ ,  $p = 0.0027$ . All other comparisons are non-significant. **B)** Data normalized within sample to control peptide. ANOVA (uncleaved CP3):  $F = 37.46$ ,  $p = 0.0076$ . ANOVA (cleaved CP3):  $F = 3.97$ ,  $p = 0.058$ . ANOVA comparing all peptides:  $F = 2.97$ ,  $p = 0.057$ . Similar results were obtained using peak area in place of peak intensity (height) for both analyses.

The advantage of this normalization method is that only identical peptides are compared to each other, removing the variation in instrument sensitivity between different peptides. However, it does not account for the run-to-run variability of the machine. I therefore repeated the analysis by normalizing within each sample to a C-terminal ovulin peptide that is not affected by ovulin cleavage. With this normalization method, I was again able to conclude that the cleaved CP3 peptide was more abundant in the CS 1 hour sample than in the other two samples (Figure 4.6B). This method also revealed a difference in the uncleaved CP3 peptide between the two samples; it was less abundant in the CS 1 hour sample (Figure 4.6B). This is expected, as the third cleavage event should decrease the level of this peptide while increasing the level of the cleaved product. A simple ANOVA found no statistical difference between the three samples for the abundance of the cleaved CP3 peptide ( $F = 3.97$ ,  $p = 0.058$ ) and a statistical difference for the abundance of the uncleaved CP3 peptide ( $F = 37.46$ ,  $p = 0.0076$ ).

This pilot experiment confirmed that the cleaved CP3 peptide abundance as measured by SRM mass spectrometry correlates with time ASM. This experiment was set up to be a proof-of-principle experiment, and not intended to be fully quantitative. Therefore, the ANOVA results should not be interpreted too strictly. For more sensitive, quantitative data acquisition, the machine would need to be run in SRM-triggered MS/MS [237], which only records the peak intensity of pre-selected peptide precursor/fragment pairs and therefore does not include a full mass spectrum.

Based on this experiment, I would argue that normalization to the control peptide would be more reliable than normalization to peptides in a separate sample. Besides removing the worry over run-to-run variation, it also appears more sensitive in detecting differences in abundance of the uncleaved CP3 peptide. This normalization method is still not ideal. A better,

though more expensive, approach would be to spike each sample prior to chymotrypsin digestion with a known amount of peptide for the cleaved and uncleaved CP3 peptides, allowing an absolute quantification of each peptide in each sample. This would completely correct for the run-to-run variation between experiments and would compare only identical peptides within each sample.

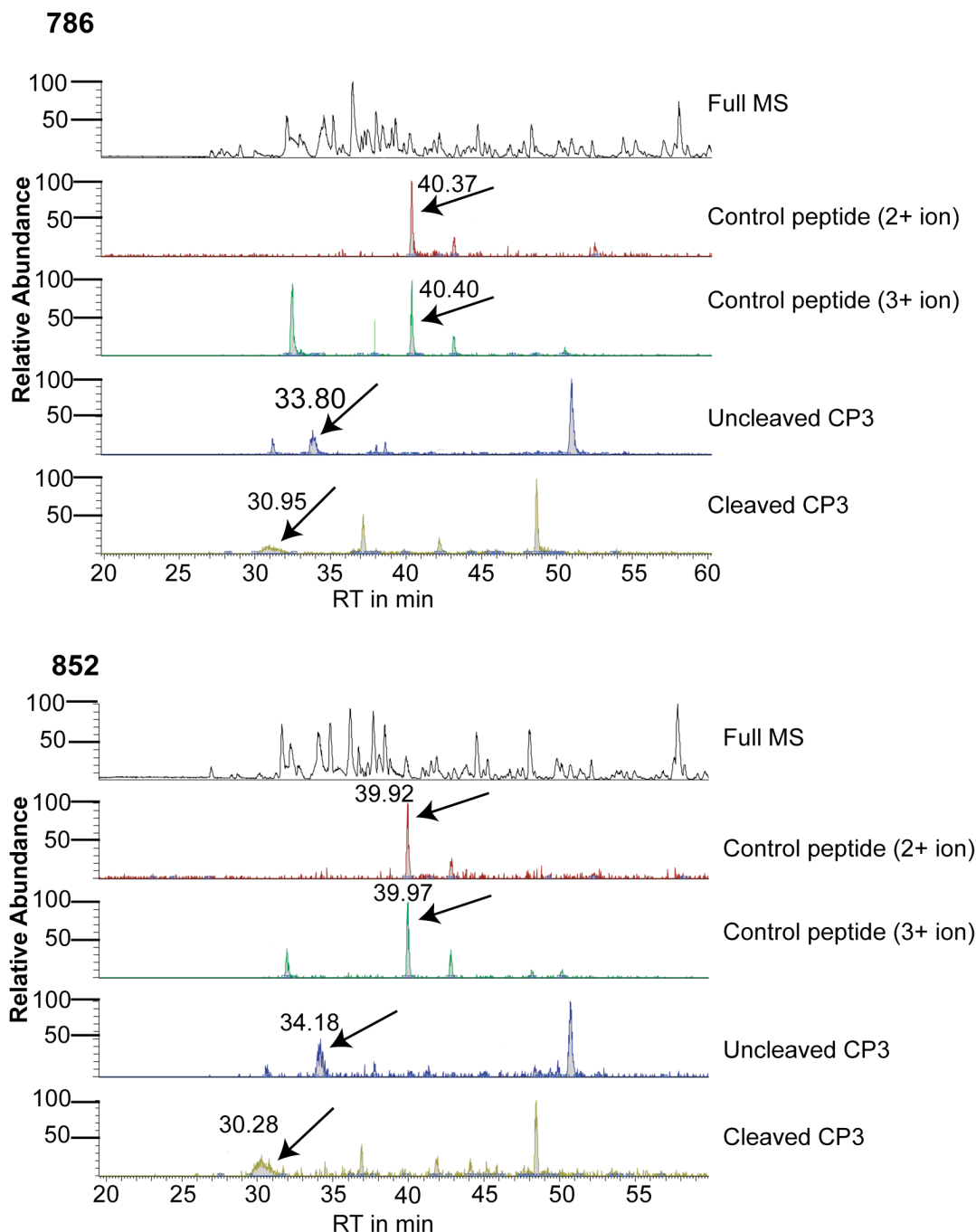
***Ovulin CP3 is not statistically different between two DGRP lines at 30' ASM***

A final pilot experiment was performed to determine if significant differences could be detected for the cleaved and uncleaved CP3 peptides between two of the DGRP lines. I chose to analyze DGRP lines 786 and 852, based on the results in my pilot quantitative Western blot study (Figure 4.4A and 3B). The earlier experiment indicated that line 786 was at an earlier processing stage at 30' ASM than was DGRP-852 (lower Q value) and should therefore have less cleaved CP3 relative to the total amount of ovulin in the sample. Again, a C-terminal ovulin peptide was used as a control for the total amount of ovulin.

The samples were run on two different instruments: the same Orbitrap instrument used in the other pilot experiments (Figure 4.7), and a more sensitive (Qtrap) instrument for quantification (Figure 4.8).

In the Orbitrap experiment, a visual inspection of the peaks suggested that there was no difference in the abundance of the control peptide between samples, but that both of the other two peptides were slightly more abundant in the DGRP-852 sample (Figure 4.8). I did not attempt to analyze these data quantitatively.



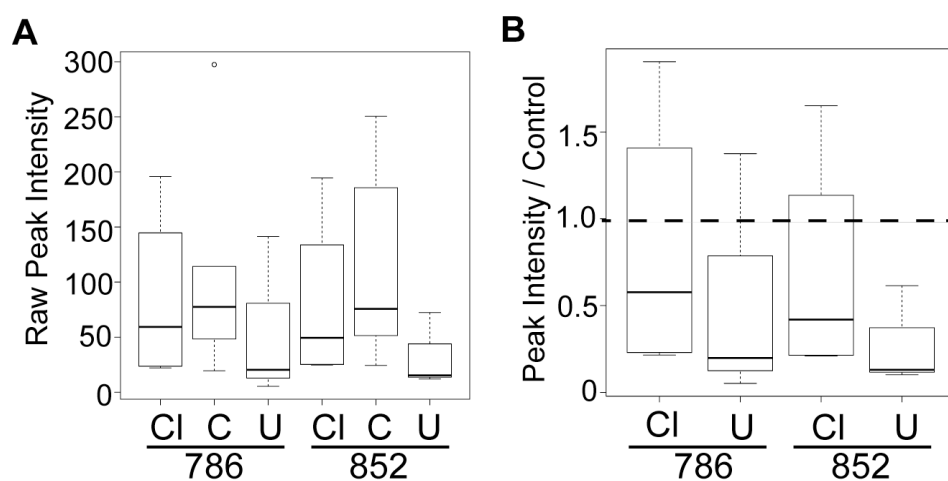


**Figure 4.7: Orbitrap data for DGRP-786 vs. DGRP-852 pilot experiment.** Peaks associated with selected peptide ions (indicated to right of spectra). **A)** Spectra from line 786. **B)** Spectra from line 852. Full MS: full mass spectra from the sample. Arrows indicate relevant peaks. The control peptide reflects the total amount of ovulin in the sample. The uncleaved CP3 peptide reflects the amount of pre-third cleavage event ovulin in the sample. The cleaved CP3 peptide reflects the amount of post-third cleavage event ovulin. Relative abundance of peptides is shown on the Y-axis. Retention time (RT) indicates time at which the selected transition ion eluted from the column.

Quantitative data were obtained from the Qtrap experiment; these are plotted in Figure 4.8A. After normalization to the control peptide, there was no significant difference between the two samples for either of the CP3 peptides (Figure 4.8B). Full results from the Qtrap experiment are shown in Table 4.3.

While I did not expect a large difference between the two lines, I did anticipate at least a trend of more cleaved CP3 in line 852 relative to 786, and the opposite trend for the uncleaved peptide. These data indicate that this assay, as used here, may not be sensitive enough for the purpose of determining the cleavage extent between DGRP lines. I propose the following steps to further optimize the SRM approach:

1. More precise normalization. After normalization to the control peptide, further normalization to the total protein abundance of the sample should be performed, to control for protein loading. The total protein abundance can be obtained by standard LC-MS/MS before running the SRM-triggered MS/MS mode.
2. Monitor more transitions for each peptide. More transitions will yield a more accurate measurement of peptide abundance.
3. Test mated females at a later time ASM. 30' ASM may be too early to reliably detect CP3. I used 30' ASM in this pilot experiment to be able to compare to the measurements previously done using Western blots at this time. However, since the CS sample at 1 hour yielded a high level of CP3, it may be a better timepoint to use.



**Figure 4.8: Qtrap data for DGRP-786 vs. DGRP-852 pilot experiment.** Data summary for experiment with peaks combined for all transition ions monitored within each peptide fragment. Relative peptide abundance is shown on the Y-axis. **A)** Raw, unnormalized data (see Methods). CI=cleaved CP3; C=control peptide; U=uncleaved CP3. **B)** Data normalized to the control peptide.

**Table 4.3: Ovulin peptides from DGRP lines 852 and 786.** Ovulin peptides from DGRP-786 or DGRP-852 females mated to males of their own line. Samples were collected at 30 min. ASM. See Table 4.2 for column explanations and peptide sequences.

Sample Name	Peptide ID	Precursor / Fragment	Peptide ion	Area	Height	Retention time
786	Control	728.8/869.6	3+	900.9	77.5	45.30
786	Control	728.8/906.6	3+	4782	297.4	45.33
786	Control	728.8/842.1	3+	2333	114.2	45.33
786	Control	728.8/970.6	3+	2222	114.2	45.34
852	Control	728.8/869.6	3+	1236	75.66	43.61
852	Control	728.8/906.6	3+	5008	250.5	43.62
852	Control	728.8/842.1	3+	2839	194.8	43.62
852	Control	728.8/970.6	3+	4201	176.5	43.65
786	Control	1092.7/869.6	2+	1574	46.52	45.33
786	Control	1092.7/1315.7	2+	834.3	50.13	45.31
786	Control	1092.7/721.0	2+	619.7	19.42	45.45
852	Control	1092.7/869.6	2+	1288	47.76	43.64
852	Control	1092.7/1315.7	2+	1302	55.07	43.64
852	Control	1092.7/721.0	2+	538.7	24.34	43.55
786	Uncleaved	754.4/1010.0	3+	6255	141.3	38.81
786	Uncleaved	754.4/958.5	3+	547.6	20.30	38.77
786	Uncleaved	754.4/1046.5	3+	241.8	5.324	38.87
852	Uncleaved	754.4/1010.0	3+	3197	72.39	37.58
852	Uncleaved	754.4/958.5	3+	581.1	15.34	37.72
852	Uncleaved	754.4/1046.5	3+	330.2	11.95	37.69
786	Cleaved	867.4/809.9	2+	9195	93.45	35.91
786	Cleaved	867.4/736.4	2+	20360	195.8	36.02
786	Cleaved	867.4/1175.6	2+	1950	25.15	35.96
786	Cleaved	867.4/1058.5	2+	1365	22.08	35.78
852	Cleaved	867.4/809.9	2+	9089	72.97	33.69
852	Cleaved	867.4/736.4	2+	21320	194.5	33.64
852	Cleaved	867.4/1175.6	2+	2103	24.72	33.64
852	Cleaved	867.4/1058.5	2+	2058	25.71	34.00

## CONCLUSIONS

I have tested three independent methods for determining the extent of variation in ovulin processing rate among the DGRP inbred lines. Using quantitative measurements of ovulin band intensities and qualitative analysis of ovulin Western blots, I have determined that there is variation, likely genetic, for this trait in the DGRP population. My qualitative Western blot data

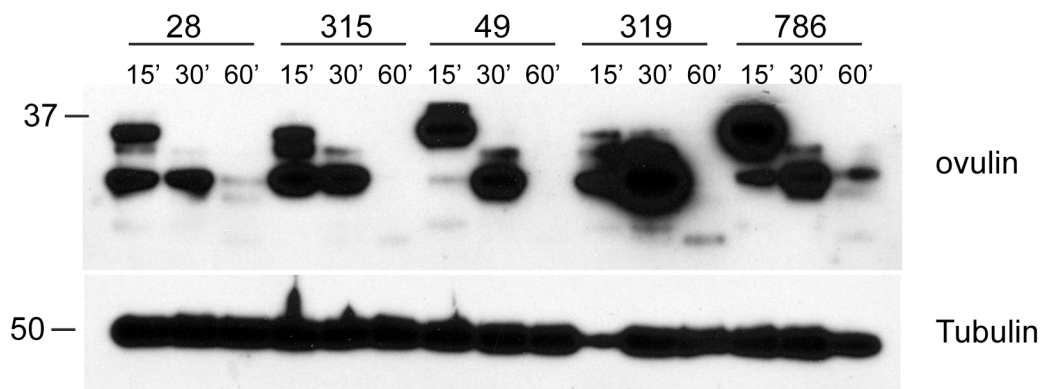
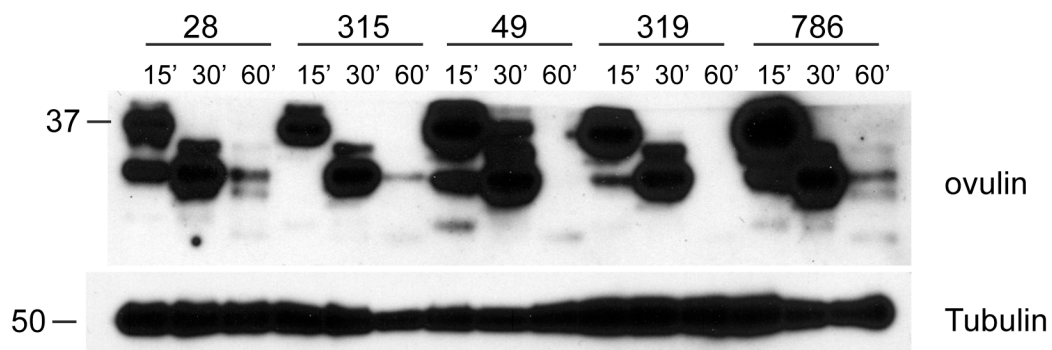
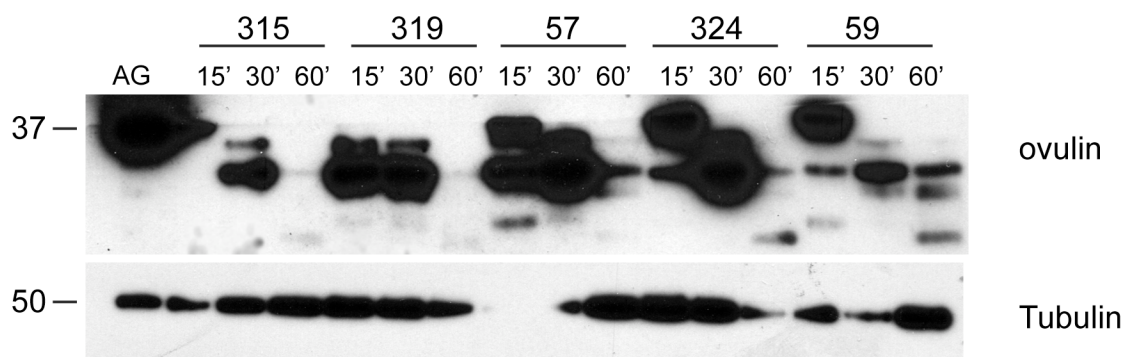
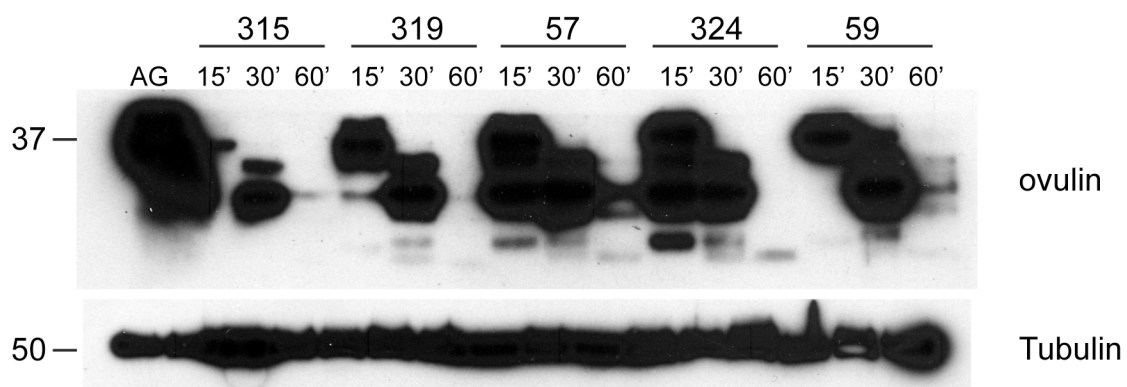
suggest that the rate-limiting step of ovulin processing may differ between lines. Further study should be done to determine what factors contribute to this variation. These factors may include variation in protease cleaving ability, ovulin structure or post-translational modifications (e.g., glycosylation), female RT pH, or variation in proteolysis co-factors contributed by the male, female, or both.

While these experiments successfully demonstrated that ovulin processing rates do vary among these lines, but are consistent between individuals of the same line, the data are not of very high resolution. Future optimization of these methods and separation of male and female effects may increase data resolution. Nevertheless, this approach as outlined here could serve as a screen for mutations of large effect.

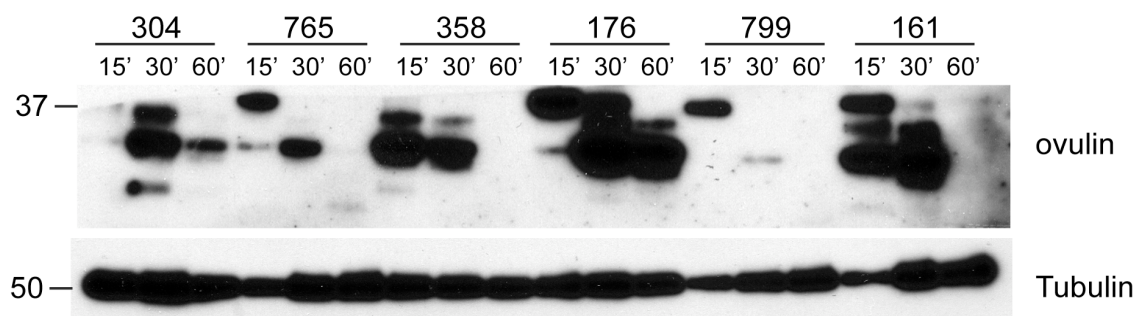
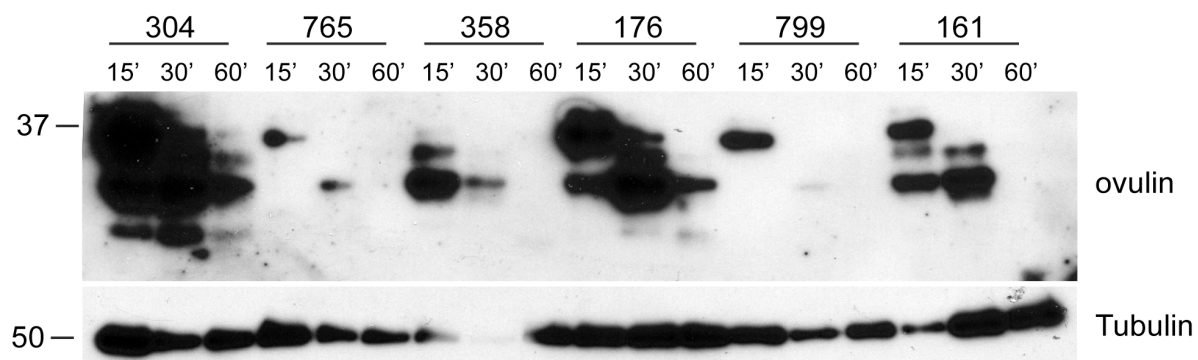
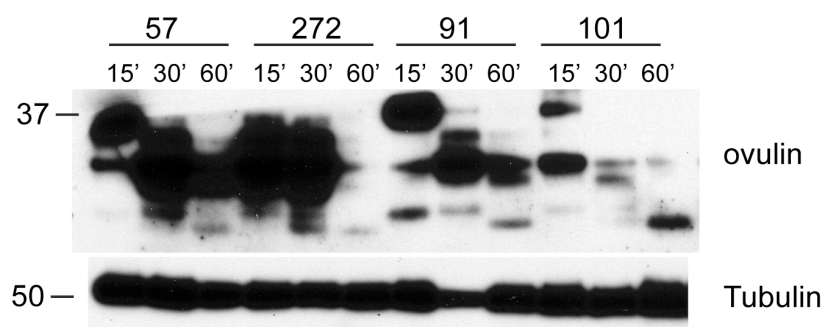
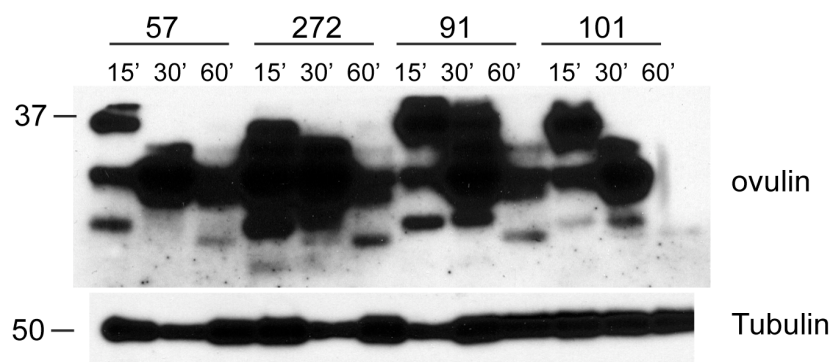
I also carried out several pilot experiments to develop an ovulin processing rate assay using SRM mass spectrometry. These pilot experiments demonstrated the suitability of ovulin cleavage product peptides for SRM analysis. This assay shows significant promise, but further optimization is required.

**Figure 4.9: Blots used for categorical ovulin processing rate study.**

All Western blots analyzed for rate categorization (See Figure 4.3 and Table 4.1). Ovulin and Tubulin Western blots are shown for each sample.

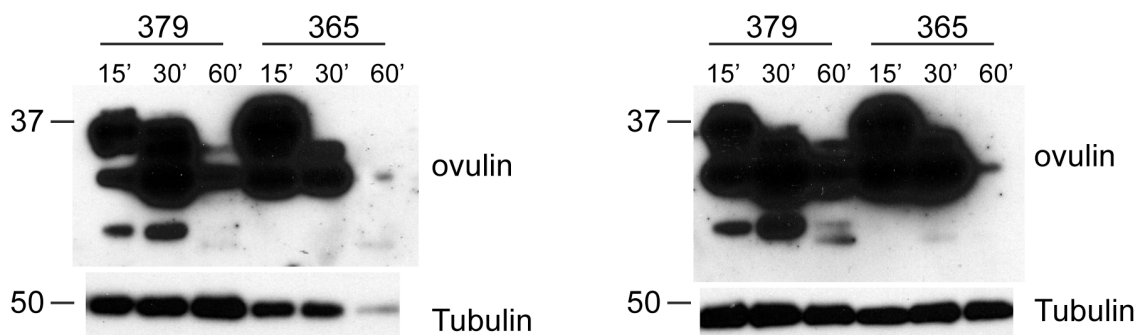
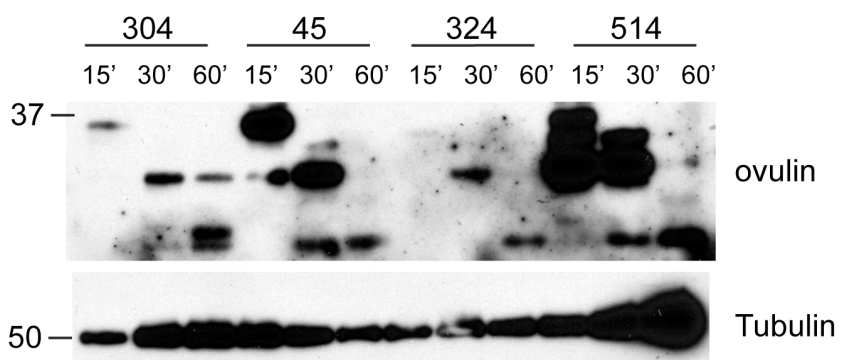
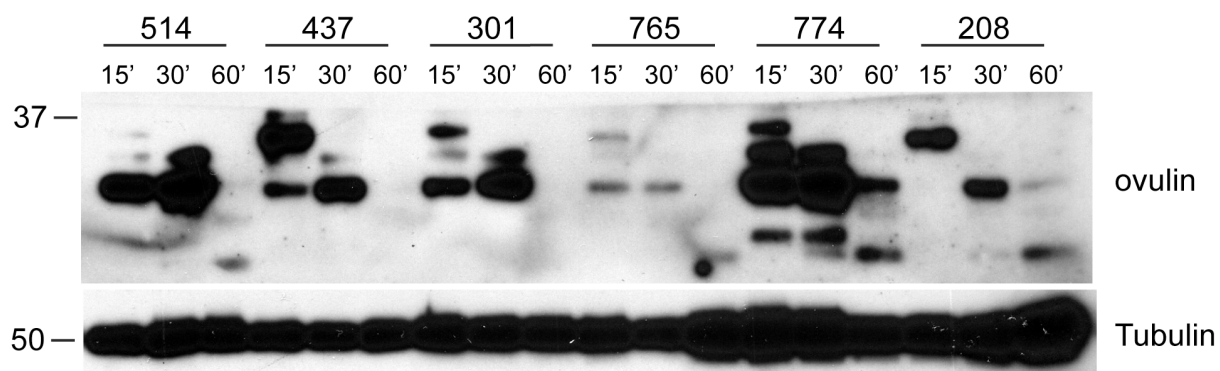
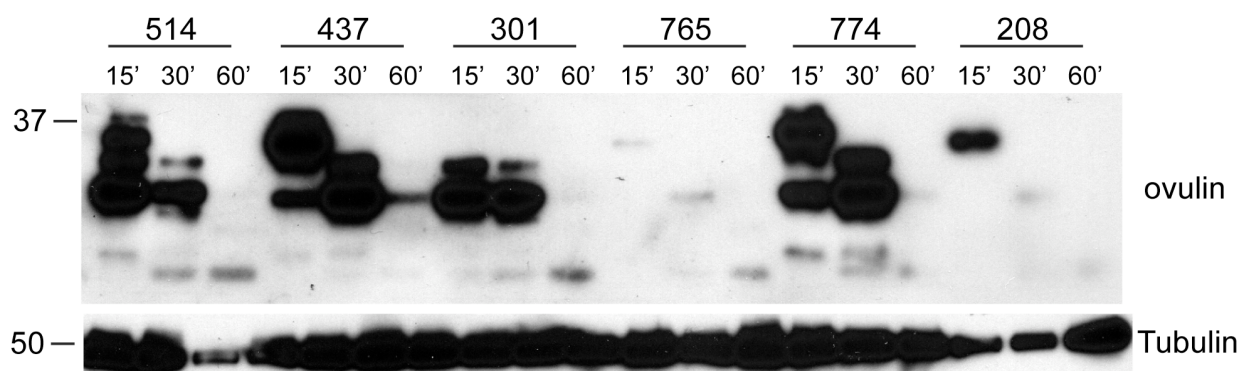


le.



**Figure 4.9 (continued)**





**Figure 4.9 (continued)**

## CHAPTER 5

### FUNCTIONAL GENOME ANNOTATION OF *DROSOPHILA* SEMINAL FLUID PROTEINS USING TRANSCRIPTIONAL GENETIC NETWORKS<sup>5</sup>

#### INTRODUCTION

The diminishing cost of high-throughput technologies such as whole genome transcript profiling, high-density genotyping and whole genome re-sequencing has shifted the focus of genomic sciences from data production to data interpretation. Foremost among the challenges in interpretation is functional gene annotation, through experimental validation or computational prediction. Even for the best-annotated genomes, a significant proportion of genes have yet to be functionally characterized [239, 240]; less than half in *Drosophila* [240].

Most knowledge regarding gene function in eukaryotes comes from mutagenesis, single gene knock-outs and RNAi knock-down experiments performed in yeasts, *Drosophila*, *C. elegans*, mouse and *Arabidopsis* [241]. These approaches have provided functions for a large number of genes in many organisms and the basis for making gene function predictions based on gene sequence similarities. However, screening large mutant collections for quantitative phenotypes is highly laborious. Furthermore, unique mutations in the same gene, or the same

---

<sup>5</sup> A version of this chapter was previously published as an article with the same title and is reprinted here with permission: Ayroles, J. F., LaFlamme, B. A., *et al* (2011) *Genet. Res. (Camb)* 93(6): 387-395. J.F. Ayroles conducted the statistical analysis of the microarray data and developed the online database, with assistance from E. A. Stone. B.A. LaFlamme conducted the experiments for Figures 5.2 and 5.3. J.F. Ayroles and B.A. LaFlamme chose the SFP score cutoff and specific genes for validation experiments.

mutation in multiple genetic backgrounds can give different phenotypes, further complicating the interpretation of such screens [242-244].

Computational methods for gene annotation complement experimental approaches. Computational methods rely on the detection of particular sequence motifs (*e.g.*, a binding domain) [245]; strong orthology with a gene of known function in a closely related species; or “guilt-by-association” [246]. The last approach is based on correlative evidence, such as the co-regulation of gene expression or the existence of known protein-protein interactions. In all cases, the functional annotation of a known gene is transferred to its interacting or correlated partner, providing a hypothesis that can be verified experimentally.

Traditionally, guilt-by-association annotation has been used in the context of environmental perturbations [247-250]. A complementary approach is to utilize natural variation in genetically correlated transcriptional networks to identify co-regulated transcripts. Previously, we used genome wide transcript profiles from 40 lines from the *Drosophila* Genetic Reference Panel (DGRP [74]), a set of inbred lines recently derived from the wild, as a source of genetic variation in gene expression. The genetic variation among these inbred lines greatly exceeds that which can be obtained by mutagenesis screens or standard genetic crosses, while sampling multiple genetically identical individuals from each line reduces environmental variance. The genetically variable transcripts are highly correlated among the lines, forming 241 transcriptional co-expression modules [74]. These co-expression modules were enriched for common Gene Ontology categories, expression in the same tissues, common transcriptional factor binding sites, and associations of gene expression with the same quantitative traits. These observations suggest that genetic correlation of gene expression with a co-expression module may be due to co-regulation and that transcripts genetically correlated with a target gene of known function are

plausibly involved in the same biological process or molecular function as the target gene [74, 251]. Here, we test this hypothesis using seminal fluid proteins (SFPs) as the focal genes.

We chose SFPs as focal genes for two reasons. First, many of the gene products of the secretory tissues of the male reproductive tract that produce the SFPs are well-understood in *D. melanogaster* [226]. This is especially true for the male accessory glands, which produce proteins known collectively as ACcessory gland Proteins (ACPs). ACPs are transferred to females in the seminal fluid and affect a number of post-mating processes [226], including sperm storage and maintenance [68, 252, 253, Ravi Ram, 2007 #412], egg production and mating receptivity [20, 203, 204], female feeding behavior [254], and sleep patterns [255]. Proteomic [57, 58] and gene expression [26] studies have identified 187 SFPs, most of which are ACPs. Second, we observed strong genetic correlations in expression among the known ACPs [74], suggesting that new SFPs, and potentially genes important for the production or function of these proteins, could be found by analyzing the correlation structure between genetically variable transcripts.

Using the DGRP expression data [74], we identified transcripts whose expression patterns correlated with known SFPs. These correlated transcripts are candidates for both previously unknown SFPs and genes that are required for regulation of SFP production. Very little is known about how SFP genes are regulated in the male; this method provides a means to identify candidate regulatory genes for further study. As a proof of principle, the only known transcription factor required for the expression of specific SFP genes [215] was among the candidate genes we identified. Though proteins encoded by regulatory genes would not necessarily be transferred to females during mating, and are therefore not SFPs *per se*, we refer to our set of candidate SFPs as cSFPs.

We identified 176 cSFP genes. For validation, we selected seven candidates with varying levels of correlation to known SFP genes and used quantitative real-time PCR to validate the correlation patterns. We also used RT-PCR to test the tissue of expression for these seven genes. We propose that this method can be widely applied to similar datasets, beyond the example of the SFP functional annotation we present.

## METHODS

### *Gene expression data*<sup>6</sup>

The gene expression data are from Ayroles *et al.* [74]. Whole genome expression was quantified using Affymetrix *Drosophila* 2.0 arrays for two replicate pools of 3-5 day old mated males and females for each of 40 DGRP lines. We median-centered the perfect match (PM) data and removed probes that were identified as likely single feature polymorphisms. We used the median  $\log_2$  signal intensity of the remaining PM probes in each probe set as the measure of expression. A total of 14,840 (78.9%) of the 18,767 transcripts on the array were expressed. Because we focus here on highly male biased transcripts, we only used the male gene expression data to identify genetically variable transcripts. We fitted the following model to the expression data:  $Y = L + e$ , where  $Y$  is the median  $\log_2$  signal intensity,  $L$  the line effect and  $e$  the residual. We identified 7,151 transcripts as genetically variable at  $\text{FDR} < 0.01$ .

---

<sup>6</sup> Gene expression analysis and statistical correlation analyses were performed by J. F. Ayroles (Harvard U. and Cornell U.), E. A. Stone (NCSU), and T. F. Mackay (NCSU). Identification of candidate Sfps was a collaborative effort between all authors. B. A. LaFlamme conducted experiments for cSFP validation.

The raw microarray data are deposited in the ArrayExpress database ([www.ebi.ac.uk/arrayexpress](http://www.ebi.ac.uk/arrayexpress)) under accession number E-MEXP-1594. The DGRP stocks are available from the Bloomington *Drosophila* Stock Center (Bloomington, Indiana).

### ***Candidate SFPs***

Of the 187 known SFP genes, 107 had genetically variable expression levels in the DGRP lines. We computed pairwise Pearson correlations between the 107 genetically variable SFPs and all 7,151 genetically variable transcripts. We then calculated an ‘SFP score’ for each of the 7,151 transcripts by tallying the number of significant correlations ( $p < 0.01$ ) with known SFPs, divided by 107. For a given transcript, a score of 100 indicates that it is correlated with all 107 known SFPs, and a score of 0.93 ( $1/107 * 100$ ) indicates the absence of significant correlation between the focal gene and any of the known SFP genes (*i.e.*, only showing correlation to itself). The thresholds used to compute this score are arbitrary, but this method is both simple and intuitive, and gives similar results to more sophisticated statistics such as the identification of eigengenes [256] following the construction of co-expression gene networks and using the PCA loadings to identify correlated transcripts may improve functional prediction.

In addition to the correlation structure, we used several criteria to identify transcripts as putative SFPs (proteins that are predominately or exclusively expressed in the male reproductive tract and likely to be transferred to females), or as potential regulatory genes (those that produce proteins unlikely to be transferred to females) but whose expression is also predominately limited to male reproductive tissues. We used FlyAtlas [62], a database of tissue specific expression for *Drosophila melanogaster*, to examine the tissues of expression for each gene with an SFP score of greater than 8. In addition, because SFPs are secreted proteins, we used SignalP

software (<http://www.cbs.dtu.dk/services/SignalP/>) to identify the presence of predicted signal sequences. The program calculates the probability that the input amino acid sequence contains an N-terminal secretion signal. Here, we used the signal peptide probability score given from the SignalP-HMM prediction method. Signal peptides are usually 15-30 amino acids long and contain a stereotypical pattern of charged, hydrophobic, and uncharged residues, though the amino acid sequence itself is not conserved [257]. However, not all secreted proteins contain predicted signal sequences [57], and not all proteins with secretion signals are secreted [257]. Therefore we do not exclude genes as being seminal fluid proteins or ACP candidates based solely on a low SignalP score.

### ***Experimental validation of cSFPs***

We chose seven genes identified as cSFPs for validation of the guilt-by-association results as well as further characterization. These genes have a range of SFP scores and a few have predicted biochemical functions, though none were predicted to be involved with SFP function. In addition to the seven candidates, we also included a known ACP gene (*CG9997* [26, 200]), and a known ejaculatory duct protein gene (*Dup99B* [258]), both of whose products are transferred to females, as positive controls. We expect cSFP genes, including those expressed in the ejaculatory duct or bulb, to correlate in expression with the known SFP, *CG9997*. We included *CG34422* as a negative control, given its low SFP score and wide expression pattern across tissues, including the male accessory glands, brain, eye, and hindgut. This gene should not show a significant correlation to *CG9997* in the quantitative PCR experiment, in contrast to the seven cSFPs.

We independently validated the tissue-biased expression results from FlyAtlas [62] for these 10 genes. We reared Canton-S males on standard yeast-glucose medium under uncrowded conditions at ~24°C. We dissected 50-60 testes (T), accessory glands (AG), ejaculatory ducts (ED), ejaculatory bulbs (EB), and male carcasses (C; no reproductive tract). Dissected tissues were placed directly into TRIzol Reagent (Invitrogen) on ice. We collected two biological replicates for each RNA extraction.

We used quantitative real-time PCR to validate the correlation structure between the genes that had been inferred from the microarray experiment. We randomly selected 20 of the 40 DGRP lines used in the microarray study [74], and isolated total RNA from two biological replicates, each with 8-12 males of each line (3-7 days post-eclosion). We then estimated the correlation of gene expression with the known SFP, *CG9997*.

### ***RNA extractions and cDNA synthesis***

We extracted total RNA by grinding dissected tissues in 150µL TRIzol Reagent (Invitrogen), following the manufacturer's recommendations for RNA isolation, except that 0.5mL chloroform was used for every 1mL TRIzol. Total RNA was treated with DNaseI (Invitrogen) and converted to cDNA with Superscript II Reverse Transcriptase (Invitrogen) and oligo-dT primers as recommended by the manufacturer. We used 500ng of total RNA per 20µl reverse transcription reaction. Negative controls without reverse transcriptase were tested once for all genes and all cDNA samples to exclude potential genomic DNA contamination.

### ***Quantitative Real-Time PCR***

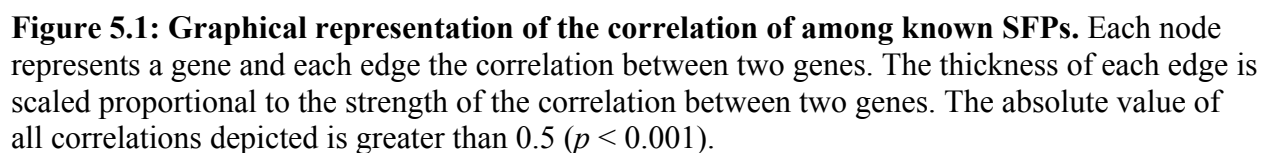


We quantified mRNA levels by quantitative RT-PCR in 25 $\mu$ L reactions with the SYBR green detection method (iQ SYBR Green Supermix, Bio-Rad) according to the protocol from MyiQ Single-Color Real-Time PCR Detection System (Bio-Rad). Each reaction was performed with 2 picograms of total cDNA, using a BioRad MyiQ Single-Color Real-Time PCR Detection system. We used the *actin5C* gene [259] as an internal standard. We used Primer3 (<http://frodo.wi.mit.edu/primer3/>) to design transcript-specific primers to amplify 85 to 148-bp regions of the genes of interest. *CG34422* primers were designed to encompass the common regions of alternative transcripts. The starting template concentration of each transcript was calculated from the standard curve of that primer pair according to the method described by Qiagen ([http://www1.qiagen.com/literature/brochures/pcr/qt/1037490\\_ag\\_pcr\\_0206\\_int\\_lr.pdf](http://www1.qiagen.com/literature/brochures/pcr/qt/1037490_ag_pcr_0206_int_lr.pdf)). We used the linear regression model  $Y = mX + b$  to quantify transcript abundance, where  $Y$  is the critical threshold (Ct) values from the qRT-PCR experiment,  $m$  is the slope and  $b$  is the intercept of the standard curve, and  $X$  is the transcript abundance. We standardized this estimate by dividing by the transcript abundance of *actin5C* in the same sample.

### ***Gene Ontology analysis***

We used Gene Ontology (GO) analysis to assign functional categories to the candidate SFP genes tested. We computed the genetic correlations between each of the seven new focal genes with the remainder of the genetically variable transcriptome. We then performed a GO enrichment analysis for the genes most strongly correlated to the focal gene ( $p < 0.001$  and  $|r| > 0.5$ .) The conclusions regarding enrichment were the same if the threshold was increased to  $p < 0.0001$ . We performed this analysis using DAVID 6.7 [260].

Of the 187 known SFPs, 107 had genetically variable transcripts among the 40 DGRP lines [74]. The 107 known SFPs were highly genetically correlated (Figure 5.1), reinforcing the idea that gene co-expression may be a reflection of shared function. We attempted to cluster this correlation matrix further into modules using various clustering algorithms, including MMC [261], but did not find strong community structure in the graph resulting from this correlation matrix. In addition, we did not find evidence supporting the idea that genes sharing a similar GO term were more strongly correlated with each other than they were to the rest of the genes.



We then analyzed the correlation matrix between the 107 known seminal fluid proteins and 7,151 transcripts that were genetically variable in males. We assigned an SFP score to each of the genetically variable transcripts based on the number of significant correlations with known SFPs (Supplementary Table 1, available online at <http://journals.cambridge.org>). We next asked whether this approach would allow us to recover the known seminal fluid proteins. We ranked the vector of SFP scores from the highest to the lowest and applied the filter that cSFPs should be expressed in male reproductive tissues based on FlyAtlas data [62]. We found that 78% of the known SFPs are in the top 500 transcripts.

We identified 176 cSFP genes that have correlated expression patterns to at least seven of the 107 genetically variable known seminal protein genes and are expressed in male reproductive tissues (Supplementary Table 1, available online at <http://journals.cambridge.org>). A total of 37 of the 176 candidates have no known or predicted functions or GO terms. An additional 13 transcripts correspond to probe sets on the Affymetrix array but not annotated genes, and could correspond to new genes. Independent confirmation of cSFP identification comes from a proteomic screen aimed at identifying male proteins transferred during mating [57, 58]. Two candidate transcripts were confirmed as *bona fide* SFPs: *CG34002* (with an SFP score of 15) and *Sfp26Ad* (with an SFP score of 41). *Sfp26Ad* was not annotated as a gene at the time we performed this experiment and corresponded to probeset 637742\_at on the Affymetrix array.

We chose seven candidate SFP genes (*CG9720*, *CG11828*, *CG31413*, *CG31493*, *CG31496*, *CG32985*, *CG34002*), as well as two positive control genes (the ACP gene *CG9997* and the ejaculatory duct protein gene *Dup99B*) and one negative control gene (*CG34422*, with an SFP score of 0.93) for validation of the microarray correlation results using real-time quantitative

PCR in 20 of the DGRP lines. The candidate genes have SFP scores ranging from moderately low (8) to very high (42, the highest SFP score found) (Table 5.1).

The real-time PCR results confirmed the correlation between all seven candidate SFPs and the known ACP gene *CG9997* across the twenty lines (Figure 5.2). As predicted, expression of the negative control *CG34422* was not genetically correlated with that of *CG9997*. However, expression of the ejaculatory duct protein gene *Dup99B*, whose gene product is transferred with the seminal fluid to females, was genetically correlated with *CG9997*, demonstrating that non-ACP SFPs can also be identified with this method.

**Table 5.1. Genes selected for experimental validation.** The SFP score is the fraction of known SFPs with which the gene had correlated expression. Sprob is the predicted probability of a secretion signal sequence as given by SignalP. Tissue of expression is given from the FlyAtlas compilation and our RT-PCR data from the male reproductive tract and carcass. Ejaculatory duct and bulb are not represented in FlyAtlas. Bold font denotes tissues of predominant expression. AG: accessory glands; ED: ejaculatory duct; EB: ejaculatory bulb; T: testis; LSG: larval salivary glands; HT: heart; HD: head. ED and EB are not represented in FlyAtlas.

Category	Gene	Affymetrix ID	SFP Score	Sprob	Tissue (FlyAtlas)	Tissue (RT-PCR)
Candidate SFA	<i>CG9720</i>	1624902_at	35	0.997	AG	<b>AG</b> , ED, EB
Candidate SFA	<i>CG11828</i>	1633604_at	41	0	AG	<b>AG</b> , ED
Candidate SFA	<i>CG31413</i>	1635084_at	42	0.987	AG	<b>AG</b> , ED
Candidate SFA	<i>CG31493</i>	1640609_at	36	0	AG	AG
Candidate SFA	<i>CG31496</i>	1628103_at	8	0.721	AG, LSG	<b>AG</b> , ED, EB
Candidate SFA	<i>CG32985</i>	1632491_at	38	0	AG	<b>AG</b> , ED, EB
Candidate SFA	<i>CG34002</i>	1625512_s_at	15	0.991	AG	AG
ACP Positive Control	<i>CG9997</i>	1634224_at	39	0.999	AG	AG
ED Positive Control	<i>Dup99B</i>	1639365_at	29	0.98	AG	AG, <b>ED</b> , <b>EB</b>
ACP Negative Control	<i>CG34422</i>	1641329_at	1	0	All but T, HT, HD	All

Table 5.1 gives SFP scores, secretion signal peptide probability, and tissue of expression for these seven genes and for the positive and negative controls. Three genes with high SFP scores were not predicted to have secretion signals. These genes' products may be secreted nevertheless, as has been seen in other cases [57], or they may be non-SFP genes that are important for the regulation of other SFPs.

Among the seven genes, all that were predicted to be expressed in accessory glands [62] were confirmed as expressed in that tissue (Table 5.1, Figure 5.3). This transcript was only seen in the ejaculatory duct, with possible low expression in the ejaculatory bulb. A possible explanation is that the ejaculatory ducts and bulbs were not examined in the FlyAtlas compendium [62], and some ejaculatory duct might have remained partially or completely attached to the accessory glands during the tissue dissections used for FlyAtlas.

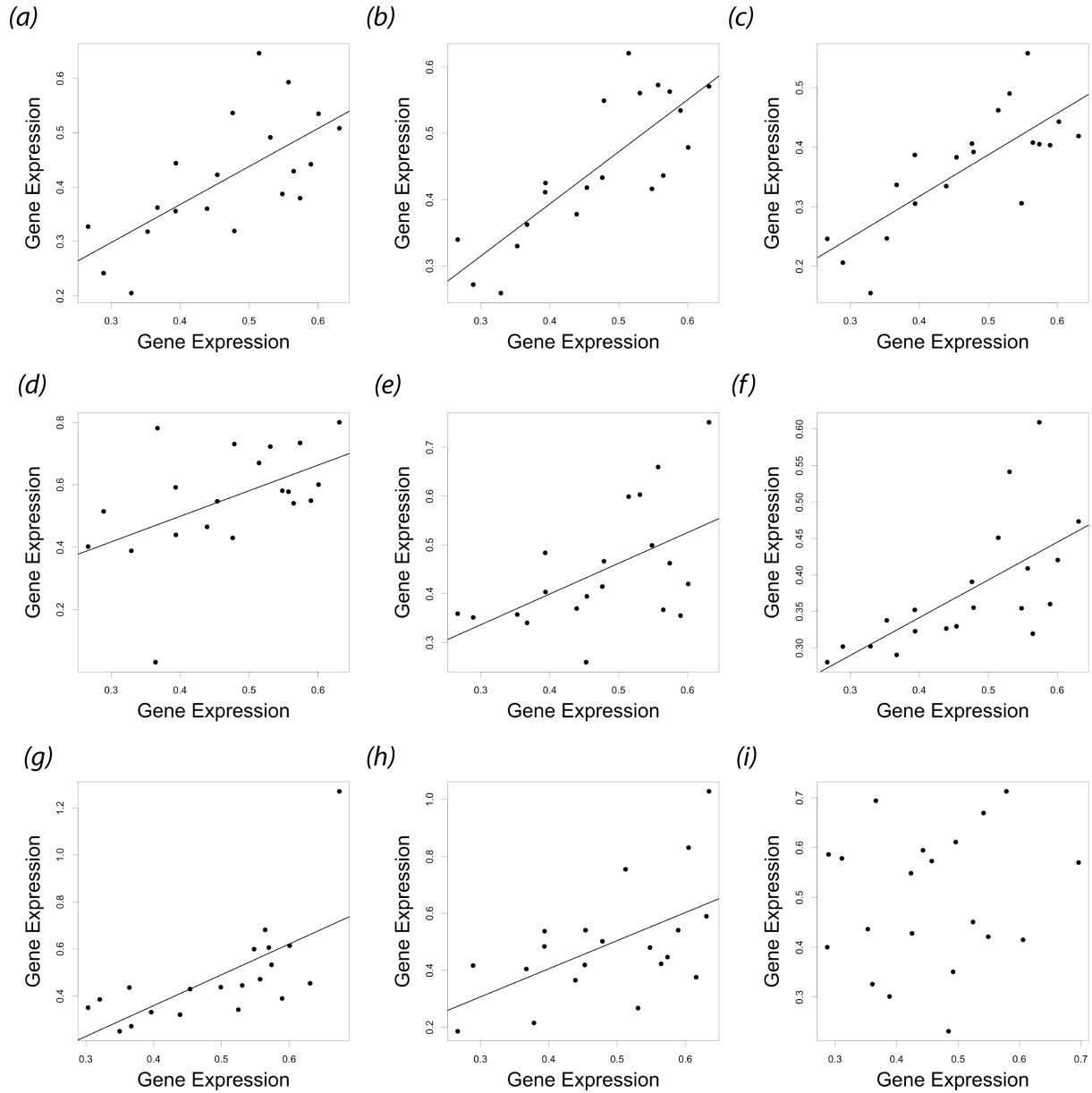
To gain insight into the possible biological processes and molecular functions of the seven candidate genes chosen for validation, we used a GO enrichment analysis implemented in DAVID (6.7) [260]. For each candidate gene, we analyzed the function of its most correlated transcripts ( $p < 0.001$  and  $r > 0.5$ ). Four of the seven candidate genes (*CG11828*, *CG31413*, *CG31493* and *CG34002*) were significantly associated with serine-type endopeptidase inhibitor activity, a predicted function shared by several other SFPs [226]. However, it is important to note that *CG11828*, *CG31413*, and *CG31493* do not contain conserved protease domains but do contain other types of predicted conserved domains. No significant GO-class enrichment was observed for *CG9720*, *CG31496* or *CG32985*.

It is possible that some of the cSFP genes are important for SFP expression and function but may not encode proteins that are transferred to females as part of the seminal fluid. As proof

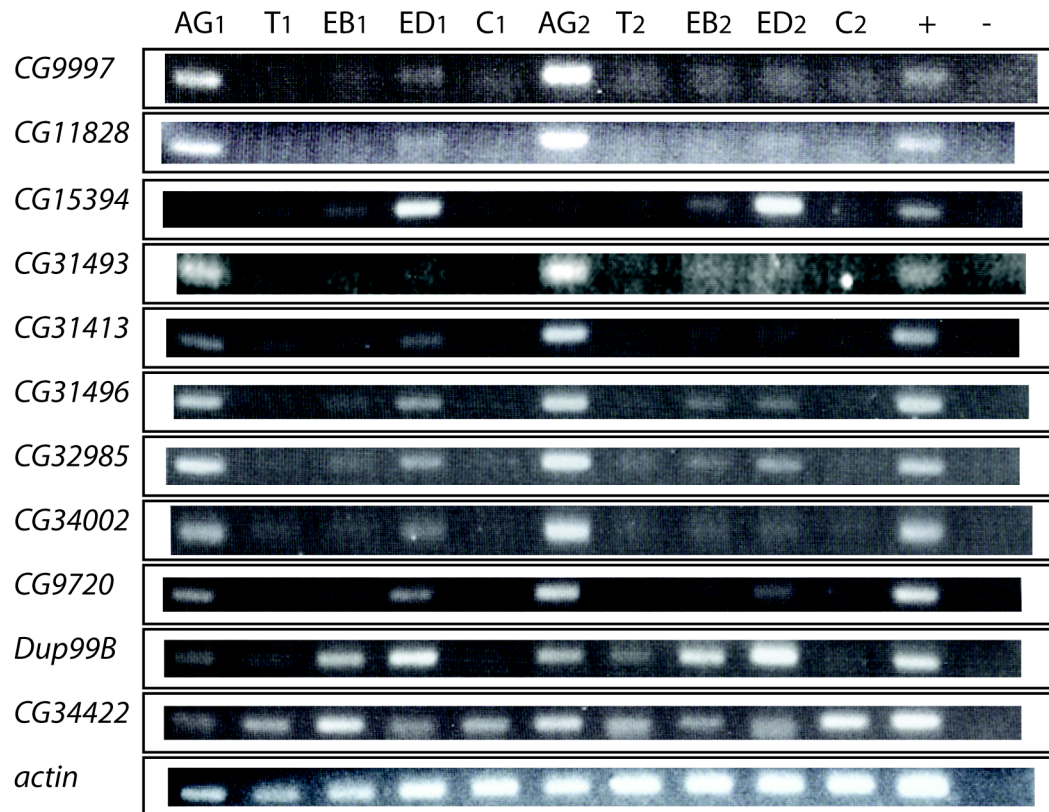
of principle that such genes can be identified by this method, our analysis detected *paired* (SFP score = 16), which encodes a transcription factor important in accessory gland development and ACP expression (Supplementary Table 2, available online at <http://journals.cambridge.org>). This *Pax* gene has a dual function in *Drosophila*: it acts first as a pair-rule gene in early embryo development [262, 263] and later is required for viability and male fertility [264-266].

Accessory gland formation and expression of at least two seminal fluid proteins expressed in the accessory gland (*ACP26Aa* and *SP*) both require the function of *paired* [215, 266].

Guilt-by-association methods most frequently rely on clustering algorithms to identify the functional membership of a candidate gene or transcript [248, 250, 267, 268]. In its most common use, guilt-by-association is used to assign functions to any or all unannotated genes that respond to a given treatment or are differentially regulated under disease conditions. Here, we have demonstrated the use of guilt-by-association methods in another context: to identify genes in a specific functional class using correlated genetic variation in gene expression among wild-derived inbred lines. This method removes the requirement for relying on arbitrary clustering or reliance on gene ontology (GO) terms to assign candidate functions to new genes. Instead, a group of genes that has been annotated and functionally clustered experimentally is used to find correlated transcripts that can then be included in the group. In this case, we used SFPs, a group defined by a biological phenomenon rather than a biochemical function. As in potentially many other cases, for example identifying genes involved in specific behaviors, GO terms do not define our selected group of genes as belonging to a biologically significant group. The group of genes we identified (cSFPs) have diverse GO functions (ranging from proteases to prohormones). A given cSFP gene could not be predicted as an SFP on the basis of GO membership.



**Figure 5.2: Gene expression level correlation between cSFPs and CG9997.** Correlation of quantitative RT-PCR estimates of gene expression between candidate SFP genes and positive and negative SFP control genes ( $y$ -axis) to a known ACP gene (*CG9997*,  $x$ -axis) among males of 20 inbred lines. All estimates of gene expression are normalized to that of *actin5C*. The linear regression line is shown, along with the  $t$ -test  $p$ -value and the estimate of the correlation coefficient,  $r$ . (a) *CG11828*,  $r = 0.68$ ,  $p = 0.001$ . (b) *CG31413*,  $r = 0.81$ ,  $p = 0.000014$ . (c) *CG31493*,  $r = 0.77$ ,  $p = 0.000063$ . (d) *CG31496*,  $r = .51$ ,  $p = 0.022$ . (e) *CG32985*,  $r = 0.53$ ,  $p = 0.015$ . (f) *CG34002*,  $r = 0.66$ ,  $p = 0.0017$ . (g) *CG9720*,  $r = 0.66$ ,  $p = 0.0016$ . (h) *Dup99B* (positive control),  $r = 0.55$ ,  $p = 0.012$ . (i) *CG34422* (negative control),  $r = 0.12$ ,  $p = 0.61$ .



**Figure 5.3: RT-PCR analysis of candidate gene expression.** RT-PCR results for candidate SFP genes and positive and negative SFP control genes in five male tissues: accessory glands (AG), testes (T), ejaculatory bulb (EB), ejaculatory duct (ED), and carcass (C, non-reproductive tissues). The subscripts denote the two biological replicates of each tissue. The number of PCR cycles for each gene was normalized to give non-saturation results. *Actin5C* was used as a control for cDNA synthesis. Whole male cDNA was used as a positive (+) PCR control, and no DNA template was used as a negative (-) PCR control.



SFP genes are well suited for this study since their expression is specific, or highly biased, to the male reproductive tract, facilitating their confirmation as SFPs; and expression of the known SFPs is genetically variable in the population of lines surveyed. An increasing number of studies are taking advantage of natural genetic variation to better understand the genetic basis of phenotypic variation [243]. In the future, the availability of sequence information for the *D. melanogaster* population used in this study will allow us to associate co-expression with eQTL analysis [243]. This additional layer of information will further our understanding of what genetic factors are driving co-expression between SFP genes, and may lead us to rethink what information should be considered when annotating a segment of sequence.

To complement this study, and generalize the simple analysis presented in this manuscript, we have created a webtool (<http://dgrp.gnets.ncsu.edu/>) that allows user to input the Affymetrix Drosophila 2.0 ID of any focal gene of interest and retrieve a vector of genes, their ranked correlation with the focal gene, as well as the gene ontology of the correlated transcripts. This tool integrates FlyAtlas information [62], allowing users to restrict the computation of correlations to genes expressed in specific tissue or to genes with strong tissue-biased expression.

Many studies using natural genetic variation to study phenotypic variation also investigate variation in gene expression and gene co-expression [243]. However, very rarely is this information translated in the form of hypothetical functional annotation for any unannotated genes involved. We advocate that such datasets be used more routinely as patterns should emerge across studies and this information will greatly improve our understanding of genes, their function, and regulation. In particular, directed analyses such as the one presented here, in which genes involved in an experimentally-defined group are sought, may help to uncover pleiotropy

among previously annotated genes and increase our understanding of how various biological systems function together.

## **ACKNOWLEDGEMENTS**

This work was funded by National Institutes of Health grants R01 GM45146 (T. F. C. M.) and R01 HD038921 (M. F. W.). B. A. L. was supported by an NSF Predoctoral Fellowship. We thank G. Findlay and E. Kelleher for comments on the manuscript, and J. Mezey and MBG colleagues for use of the qPCR machine. This is a publication of the W. M. Keck Center for Behavioral Biology.

### **Supplementary Tables (available online at <http://journals.cambridge.org>)**

- Supplementary Table 5.1. Genes for which variation in gene expression among the DGRP lines was correlated with at least one known SFP, and corresponding SFP scores.
- Supplementary Table 2. Candidate SFP genes with correlated expression patterns to at least seven known SFPs, and which are expressed in male reproductive tissues.

## CHAPTER 6

### DISCUSSION AND FUTURE DIRECTIONS

#### *Significance of this thesis*

Identifying the functions of individual seminal fluid proteins (Sfps) is a significant challenge, given the molecular complexity of seminal fluid. In addition to non-protein components, seminal fluid contains hundreds of proteins with a wide range of molecular weights and biochemical properties. While seminal fluid as a whole has many functions, a single Sfp may have only one function that is difficult to detect within the background of this complexity. In addition, many seminal fluid proteins are likely to have redundant functions, or functions that are only detected under specific conditions. Furthermore, while some Sfps may have large effects on fertility (e.g., sex peptide [203, 204] in *Drosophila melanogaster*), others may have much smaller effects (e.g., ovulin [105], also in *D. melanogaster*) that have the potential to be significant for males in the wild, but not easily detectable using standard laboratory assays. Therefore, single knock-out experiments may not be successful in characterizing the functions of all Sfps in any animal, including *Drosophila*. One way to meet the challenge of identifying functions of large numbers of Sfps is to focus our efforts on individual Sfps that are likely to play key roles in regulating Sfp-mediated pathways, such as proteolysis regulators, and use the results of those single knock-out experiments to understand the roles of the downstream proteins in the pathway.

In this thesis, I have undertaken the dissection of one such pathway in *Drosophila melanogaster* seminal fluid. First, I identified the predicted serine protease ‘seminase’ (CG10586) as a key regulator of two, apparently otherwise independent, Sfp pathways [Chapter

2]. One pathway coordinates the localization of three other Sfps (CG9997, CG1656, CG1652) to the female seminal receptacle. Proper localization of these proteins is required for sex peptide to bind sperm, from which it can be gradually cleaved and released in females to induce long-term post-mating responses, such as elevated egg-laying and reduced receptivity [67, 68]. Seminase itself localizes to the seminal receptacle, where it may physically interact with these other Sfps. RNAi knockdown of seminase in the male leads to severe fertility defects: females mated to these males lay many fewer eggs over a 10 day period than females mated to control males. Even more detrimental for the male, these females will re-mate much sooner to other males. Both of these effects (reduced egg-laying and earlier re-mating) can be attributed to the failure of sex peptide to properly bind to sperm. The second pathway which seminase regulates leads to the activation of the predicted astacin metalloprotease CG11864. CG11864 cleaves two downstream Sfps in the female: ovulin and Acp36DE [81]. Thus, seminase regulates at least one protease cascade in *Drosophila* seminal fluid; whether it functions as a protease in the sex peptide pathway is not yet known.

The characterization of seminase presented in Chapter 2 is a significant step forward in the study of Sfps. I have identified an Sfp protease as a master regulator of at least two Sfp-mediated pathways that begin in the male and are completed in the female. This study demonstrates the importance of Sfp proteolysis as regulators of pathways important for fertility; further understanding of the regulation of seminase function will be essential in understanding how Sfp-mediated pathways are regulated, how they evolved, and the consequences of mis-regulation of proteases in the seminal fluid. The seminase-regulated post-mating pathways will provide a useful model system for understanding the role of proteolysis in seminal fluid that can be applied to studies in other organisms.

The mechanistic studies presented in Chapter 3 further dissect the seminase/CG11864 proteolysis pathway by demonstrating the functional significance of a predicted activational cleavage site in CG11864 and the importance of a *Drosophila melanogaster*-specific catalytic site residue substitution in CG11864. Additionally, my work, as described in Chapter 3, has established that the male ejaculate contains the necessary proteases to cleave Acp36DE and at least partially cleave ovulin; protein-free buffers, given a long enough incubation time, may be able to compensate for the contribution of females to the proteolysis of these Sfps. On the other hand, females can express an activatable form of CG11864 that compensates for CG11864 deletion in males. The results in Chapter 3 represent an important step in defining the spatio-temporal regulation of an Sfp protease in a genetically tractable model organism. They also will be paramount in developing an *in vitro* assay for ovulin and/or Acp36DE processing.

My experiments in Chapter 4 contribute towards a characterization of the rate of ovulin processing in wild-derived inbred *Drosophila* lines and present several methods for qualitative and quantitative measurements of this rate. Classical mutagenesis experiments are not suitable for identifying additional members of the ovulin processing pathway because of the limitations on sample size and the small magnitude of differences one might expect in ovulin processing rate due to single mutations (as discussed in Chapter 4). Identifying new members of the ovulin processing pathway is a crucial part of fully characterizing the seminase/CG11864 protease cascade. The *Drosophila* Genetic Reference Panel lines [219] provide an excellent resource to identify these additional pathway members: ovulin processing rates measured in the lines can be used in future association studies to identify candidate proteins, in either males or females, that regulate the proteolysis of ovulin.

I was unable to detect any fertility consequences of removing CG11864 from the male

ejaculate (Chapters 2 and 3). This likely indicates either that loss of cleavage of ovulin and Acp36DE is compensated for by other proteins, or that the fertility consequences could not be detected using the specific assays I employed (e.g., phenotypes may be affected that I did not screen for). Additional assays should be performed that mimic situations likely encountered by flies in the wild. These include, but are not limited to effects of sperm competition, male or female age, ejaculate depletion, or nutritional status. For example, after several matings, males will be limited for both sperm and Sfps. In this situation, processing of ovulin may be more important (if it releases additional bioactive molecules) than might be the case when other Sfps (e.g.: sex peptide) are more abundant and can compensate for a loss of CG11864 activity. Temperature may also play a role in the importance of the seminarese/CG11864 proteolytic pathway. Thus far, all experiments have been conducted at standard laboratory temperatures, though flies experience a far wider range of temperatures in the wild. All of these factors could affect our ability to detect the functional consequences of loss of CG11864 from the ejaculate.

Finally, Chapter 5 presented an important complementary method to the recent proteomic studies [57, 58] (and earlier transcriptomic studies [26, 55]) undertaken to identify additional *D. melanogaster* Sfps. This gene expression profile-based identification of candidate Sfps identified new Sfps, including many new predicted proteases and protease inhibitors that warrant further study for their roles in fertility. Furthermore, this method has the potential to identify important proteins that are required for the regulation or function of Sfps, but are not Sfps themselves (e.g., transcription factors and chaperones). Future studies of these proteins may yield greater insight into the regulation of Sfp-mediated pathways, including protease cascades.

### ***Future directions***

In the summary above, I have outlined some broad directions for future work based on the results presented in this thesis. However, there are several specific experiments, which I believe are important for the short-term continuation of this work.

I described above several conditions that could be varied to test for an effect on egg-laying due to CG11864 loss. In addition to testing for an effect of these conditions (age, temperature, competition, ejaculate depletion), other phenotypes should also be examined. One possibility is to test for a structural difference to the mating plug, since Acp36DE is a known component of the mating plug. This would require the development of a new assay to determine whether there is a difference in size, density, or perhaps auto-fluorescence.

The fate of seminease in the mated female should be determined. In Chapter 2, I demonstrated that some seminease protein localizes to the seminal receptacle. With the reagents available to me at the time, I was not able to determine whether seminease localizes to any other tissues within the female (e.g., the oviduct, spermathecae, or targets outside of the female reproductive tract). I was also not able to conclusively determine whether seminease localization to the seminal receptacle is disrupted when other proteins in the sex peptide pathway (CG9997, CG1652, CG1656, or CG17575) are knocked down, due to the difficulty of detecting seminease at all in the seminal receptacle using Western blots. Unfortunately, the seminease antibody I generated in Chapter 2 is not likely to be suitable for immunohistochemistry. Therefore, I suggest generating a fly line expressing a tagged version of seminease (e.g., with GFP) to identify other targets of seminease in the mated female (assuming that the fusion tag does not interfere with seminease's function; at least two different tags should be tried to test this). The tagged seminease protein could also be used to determine whether seminease, like ovulin and Acp36DE [76], enters the female circulatory system. Fly lines could also be generated to express tagged

forms of seminare cleavage products, after these have been defined, to determine whether different regions of seminare localize to the same tissues.

The regulation of seminare in the male reproductive tract is not yet clear. Seminare is rapidly cleaved to a smaller molecular weight protein upon ejaculation. This is reminiscent of the pro-peptide cleavage activation mechanism of many proteases. A conserved domain-based prediction exists for the seminare cleavage site, as well as its catalytic and substrate-binding sites. However, testing the functional consequences of mutations at these sites, as done for CG11864 in Chapter 3, requires a seminare null. Identification or generation of a suitable null allele of seminare, as with homologous recombination, is a critical step in further dissecting its activation mechanism. Removal of the seminare pro-peptide will establish whether an active form of seminare can be expressed in the male accessory glands; if so, can CG11864 be activated to cleave its substrates in the male? Or, does the male somehow compensate to inactive an inappropriately active protease?

The same experiment (removal of the pro-peptide) also needs to be performed with CG11864. I generated a fly line expressing a pro-peptide-deleted CG11864 in the male accessory glands. Unfortunately, this protein was not secreted into the seminal fluid for transfer to females. One explanation for this result is that the predicted CG11864 secretion signal sequence is not correct. A fusion of the empirically-validated ovulin secretion signal sequence [108] onto the predicted active CG11864 protein sequence would clarify this. If this doesn't work, an alternative explanation is that I did succeed in expressing active CG11864, but that the active protease is somehow sequestered in the accessory gland to prevent detrimental proteolysis therein. Expression of a catalytically dead, pro-peptide deletion form of CG11864 could determine whether this is the case: if the presence of an active protease is the signal for



sequestration, then inactivation of the protein should allow it to be transferred to females.

The exact role of seminase in CG11864 activation could be further analyzed. The predicted CG11864 pro-peptide cleavage site is not a tryptic cleavage site, as would be predicted if seminase was the only protease necessary to cleave CG11864. As I discuss in Chapter 2, the prototype of the astacin metalloprotease family, *Astacus astacus* astacin is cleaved first by trypsin at a site (or sites) near the final pro-peptide cleavage site and second by intramolecular cleavage of astacin itself [208]. Both steps are required for proper activation of astacin. A similar mechanism may be used to activate CG11864. In support of this hypothesis, several tryptic sites exist near the predicted CG11864 pro-peptide cleavage site. These sites could be mutated, alone or in combination, to determine if the partial CG11864 cleavage seen in females mated to CG11864 cleavage mutant males still occurs. If not, this lends further support the hypothesis that CG11864 activation is a multi-step process.

Whether or not seminase is required for the activation of female-expressed CG11864 could not be established in my experiments in Chapter 3. In females ectopically expressing the wild-type CG11864 protein, CG11864 is present as two different molecular weight forms. CG11864 from females ectopically expressing the cleavage mutant CG11864 only produce the larger protein form. This suggests that wild-type female-expressed CG11864 is cleaved (and potentially activated) prior to mating and receipt of seminase. However, in my experiments all matings involving ectopic expression females were done with males that transfer seminase. Therefore, it is impossible to determine whether seminase is required from the male for ovulin cleavage when CG11864 is ectopically expressed in females. Generation of a CG11864 null, seminase knockdown (or knock-out) double mutant would answer this question. Males of this line could be mated to females ectopically expressing wild-type CG11864. If ovulin and/or

Acp36DE is still cleaved after these matings, then seminase is not required to activate female-expressed CG11864.

Finally, and possibly most importantly, an *in vitro* system for cleavage of CG11864 substrates should be developed. At the present, CG11864 and seminase are still “predicted” proteases. An *in vitro* proteolysis system, using purified CG11864, seminase, and ovulin (or Acp36DE) would demonstrate conclusively that these Sfps are proteases and that they directly act on their predicted substrates (i.e., ovulin and Acp36DE), if this is in fact the case. I present preliminary experiments carried out toward this goal in Appendix 1. More work is required to optimize the purification and concentration of all pathway components (ovulin, CG11864, and seminase). Additional work using the ejaculate collection experiments described in Chapter 3 will be important in identifying buffers to use for *in vitro* ovulin cleavage assays. Once this assay is developed, mutations can be made in pathway components to further dissect the regulation of both seminase and CG11864. A key experiment would be to determine whether CG11864 is capable of self-activation *in vitro*. Astacin can cleave itself, however this activation is much less efficient than activation with trypsin [269]. If CG11864 can catalyze its own cleavage, like astacin, a follow-up experiment would be to purify a catalytically dead CG11864 protein. Cleavage with seminase, followed by mass spectrometry, would then identify the exact residues cleaved by seminase. Addition of active CG11864 would then determine whether CG11864 cleaves intramolecularly or intermolecularly.

These experiments will lead to the further development of the seminase/CG11864 pathway as a model system in which to study seminal fluid proteolysis. Besides allowing for hypothesis-testing of predicted regulatory mechanisms of such pathways, this system will also be important for studying the evolution of seminal fluid protease cascades. The genomes of twelve

*Drosophila* species have been sequenced, allowing for evolutionary analysis of pathways regulated by *seminase* and CG11864. For example, generalities about the evolutionary rates of proteases and targets of proteolysis may be uncovered. To date, only male-derived proteins have been identified as part of the *seminase*/CG11864 proteolytic pathway. This may be simply because the pathway needs to be activated quickly, i.e.: before entering the female reproductive tract, so that reliance on female proteases would be detrimental to the system. However, there is also the possibility that this proteolysis pathway is mainly beneficial to the male. In that case, we might expect to find protease inhibitors expressed by the female that slow down the processing of ovulin, Acp36DE, or other not yet identified substrates of CG11864. There are a large number of predicted protease inhibitors in the female reproductive tract that could be tested for an interaction with the *seminase*/CG11864 pathway.

Ultimately, identifying correlations between the mating systems of these twelve *Drosophila* species and their *seminase*/CG11864 pathway components may lead to new discoveries about in the relationship between proteolysis-regulated seminal fluid pathways and particular aspects of the male-female interaction before, during, and after mating.

## APPENDIX 1

### IN VITRO EXPRESSION OF OVULIN AND PROTEOLYTIC ENZYMES

#### OVERVIEW

The predicted proteases seminease and CG11864 are expressed exclusively in the *Drosophila melanogaster* male accessory glands, and both are required for the processing of two other accessory gland proteins (Acps): ovulin and Acp36DE [63, 81]. In order to demonstrate that seminease and CG11864 act directly on their inferred substrates, it is important to reconstruct the seminease-regulated proteolysis pathway *in vitro* using purified components.

The optimization of an ovulin *in vitro* cleavage assay is necessary to determine whether CG11864 acts directly on ovulin as a protease. By removing the predicted pro-peptide of CG11864, I can test whether removal of this pro-peptide renders the protease active, as predicted [81]. Similarly, *in vitro* cleavage of CG11864 by seminease would establish that seminease acts directly on CG11864 as a protease. In the long-term, such an *in vitro* cleavage assay for CG11864 and/or ovulin would allow for direct tests of the functional significance of mutations in any of these proteins.

Here, I describe the experiments done thus far to determine the best methods for expression and purification of ovulin, CG11864, and seminease. Additional experiments are needed to optimize these methods for greater protein yield and test for *in vitro* cleavage of ovulin and CG11864.

#### METHODS AND RESULTS

##### *Cell-free expression kits*

Three cell-free expression kits were tested to determine their ability to efficiently express ovulin, CG11864, and/or seminease proteins. These different systems used were based on lysates from three origins: *E. coli*, wheat germ, and insect cells.

##### *1. Invitrogen Expressway mini*

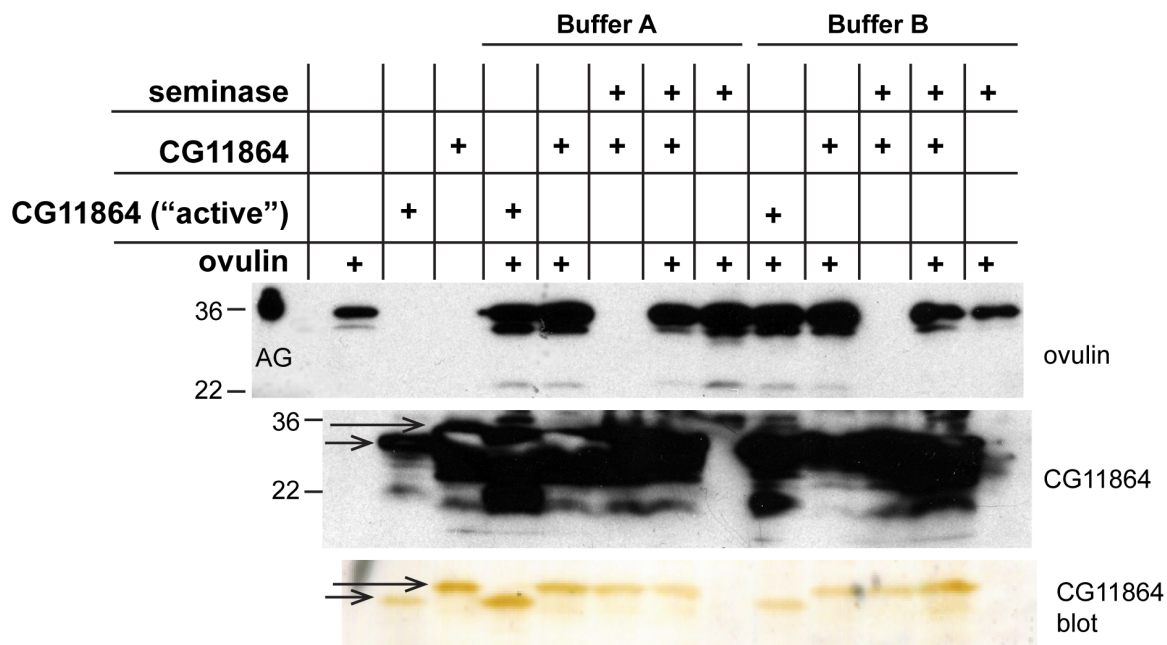
The first kit tested was an *E. coli* cell-free lysate system (Invitrogen Expressway mini). The vector pExp1-DEST (Invitrogen Gateway system) was used to express four constructs: full-length ovulin, full-length CG11864 (without signal sequence), CG11864 with signal and pro-sequences deleted (“active”), and full-length seminease. The “active” form of CG11864 is based on its predicted pro-peptide sequence, which is presumed to prevent CG11864 activity [81]. The pExp1-DEST vector includes an N-terminal His tag. See Table A1.1 for details.

Following protein expression (using the manufacturer’s protocol), I concentrated the reaction volume using centrifugal filter columns with a 3kDa molecular weight cutoff (Millipore Microcon) to approximately 10μL. A final spin was performed with one of two buffers for buffer exchange:

A: 25mM Tris-HCl, pH 9.0, 2.5μM ZnSO<sub>4</sub>, 0.005% Tween-20 [234]

B: 50mM Tris-HCl, pH 7.5, 1mM MgCl<sub>2</sub>, 1mM CaCl<sub>2</sub> [234]

I then mixed buffer-exchanged reactions in equal volumes and incubated them with shaking at room temperature for 3 hours. I separated proteins in the samples on 7-15% acrylamide gels and performed Western blots using anti-ovulin and CG11864 antibodies (the seminease antibody was not available at that time). Results from this experiment can be seen in Figure A1.1.



**Figure A1.1: Ovulin and CG11864 are expressed *in vitro* in an *E. coli* extract system.** A Western blot, probed with anti-ovulin (top panel) and anti-CG11864 antibodies. A male accessory gland sample (AG) is shown on the ovulin blot for size comparison. *In vitro* transcription/translation reactions containing the expression vectors shown above the blot were mixed and incubated at room temperature to determine if ovulin could be processed with *in vitro*-expressed enzymes. The CG11864 Western was over-exposed, making it difficult to see individual bands on the film. An image of the blot (with visible signal) is shown in the bottom panel. Numbers to the left of the blots indicate approximate sizes in kDa.

I was able to produce ovulin protein using this system (Figure A1.1, ovulin panel). Ovulin incubated on its own did not show any evidence of degradation or processing (Figure A1.1, ovulin only lane). *In vitro*-expressed ovulin runs at a slightly lower molecular weight than *in vivo*-expressed ovulin due to lack of post-translational modifications. Samples containing ovulin mixed with any of the other three reactions (full-length CG11864, “active” CG11864, or seminase) had an additional, approximately 25kDa band, which may be indicative of cleavage (Figure A1.1A). However, there appeared to be no difference between the two different forms of CG11864, and even seminase alone had the same effect. It is possible that some unknown protease in the *E. coli* extract was being activated by seminase and/or CG11864. Bands corresponding to the other known cleavage products of ovulin were never observed. In addition, the above result (with the 25kDa band) were not fully reproducible.

I was also able to get CG11864 expression with this system (Figure A1.1, CG11864 panels). There was a discernable size difference between full-length CG11864 and the “active” CG11864 (due to deletion of the pro-peptide sequence from the “active” CG11864 insert). Given the requirement for seminase in CG11864 cleavage, I expected that addition of seminase *in vitro* reaction mixture would result in a size shift of full-length CG11864 to the smaller size of the “active” product. However, there was no such effect of seminase addition (Figure A1.1, lanes with seminase added).

I was not able to determine whether seminase was expressed in the *E. coli* extract system, as there was no antibody for the protein at the time that I did these experiments. Earlier experiments using inducible expression in *E. coli* followed by Coomassie staining had determined that full-length seminase could not be expressed in *E. coli* (data not shown), though whether this would also be the case in the cell-free lysates is unclear. There is not enough protein expressed with the *in vitro* system for detection of expressed proteins by Coomassie staining.

## ***2. Eukaryotic cell-free extract kits***

The Invitrogen Expressway system described above is not ideal since it is based on *E. coli* rather than on a eukaryotic system. Potentially important post-translational modifications, such as glycosylation, will not occur in this system. I therefore tested two eukaryotic systems: a cell-free wheat germ extract system (Promega TNT Coupled Wheat Germ Extract) and an insect cell-free lysate system (Qiagen Easy Xpress Insect Kit II). Manufacturer protocols were followed for both kits.

I next put all constructs into the pExp2-DEST vector (Invitrogen Gateway system); this vector includes a C-terminal His tag. Native stop codons were kept intact to enable expression of untagged proteins. See Table A1.1 for details. Use of this vector allowed signal sequences to be included without addition of a fusion tag, so that each protein would be recognized by secretion pathway components. I was not able to get ovulin expression using this plasmid in the wheat germ extract kit (not shown). I did see ovulin expression from the insect cell-lysate kit, though at very low levels (Figure A1.2A).

CG11864 was expressed in both systems (Figure A1.2B). CG11864 from the wheat germ extract kit was expressed as a smaller molecular weight product than in accessory glands (less than 33kDa) (Figure A1.2B, lanes 1, 2 and 3) while CG11864 from the insect cell-lysate kit was expressed as a larger molecular weight product (greater than 33kDa) (Figure A1.2B, lanes 6 and 7). Smaller bands present in CG11864 expression samples using the insect cell-lysate kit were likely CG11864 degradation products (Figure A1.2B, lanes 6 and 7). Expression using both the CG11864 and ovulin plasmids (total plasmid concentration double that of single plasmid experiments) resulted in apparent loss of CG11864 expression, since these samples appeared identical to samples prepared using only the ovulin expression plasmid (Figure A1.2B, lane 4). A cross-reactive band appeared that was the same size as the “active” form of CG11864, but was also present when CG11864 plasmid was not used (Figure A1.2B, lane 5). Seminase expression was also tested (after antibody development), but was not detected using either kit (not shown).

## ***Semi-in vitro processing tests***

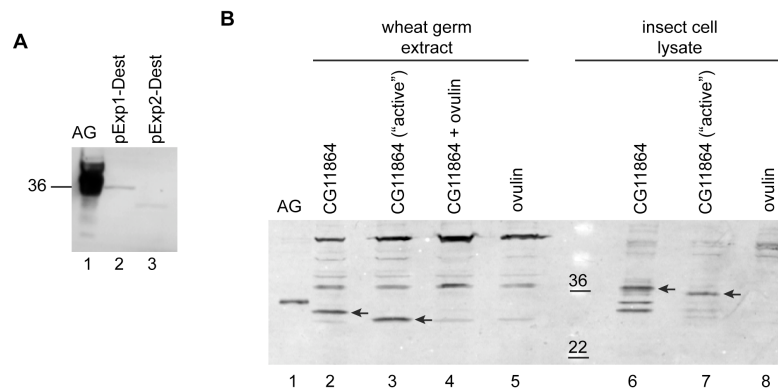
Because I could not get sufficient ovulin expression using these *in vitro* expression systems, I next attempted semi *in vitro* experiments by mixing *in vitro*-expression CG11864 reactions with extracts of mated female reproductive tracts (females were mated to CG11864 null males (see Chapter 3) and dissected within 3 hours after the start of mating). Ovulin is not normally cleaved in females mated to CG11864 null males, and as long as the ejaculate has not been expelled from the uterus, full-length ovulin is always detected in the female reproductive tract within this timeframe. Thus, *in vitro*-expressed CG11864 would be allowed to act on endogenous ovulin in the presence of other Sfps and female proteins. This experiment tested whether CG11864 expressed *in vitro* is an active protease under the conditions used.

Full-length CG11864 and “active” CG11864 were expressed using the two eukaryotic

expression kits. Each *in vitro* cleavage assay contained two female reproductive tracts after mating to CG11864 null males, 10 $\mu$ L reaction mix, containing either full-length or “active” CG11864, and 30 $\mu$ L Buffer A. Samples were incubated at room temperature with shaking for 4 hours. Following incubation, I concentrated all samples to 10-15 $\mu$ L with Microcon filters (Millipore) with a 3kDa molecular weight cutoff. Proteins in concentrated samples were separated with SDS-PAGE, transferred to PVDF membranes and probed with anti-ovulin antibody.

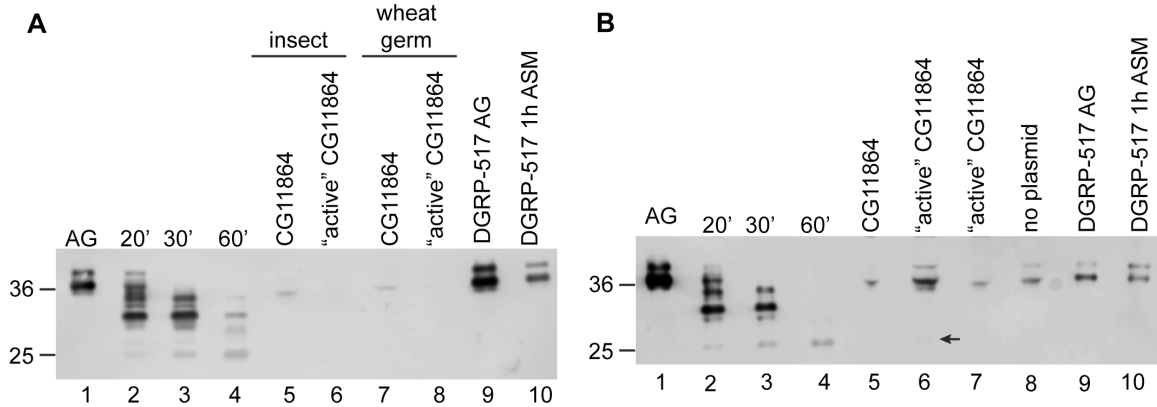
Full-length ovulin was visible in samples containing full-length CG11864 expression reactions (Figure A1.3A, lanes 5 and 7). However, I did not detect ovulin in samples that contained “active” CG11864 expression reactions (Figure A1.3A, lanes 6 and 8). Therefore, I could not determine whether ovulin processing had occurred; ovulin might not be visible because it was fully processed, or simply because the protein was somehow lost during the experiment.

I repeated the experiment using only the samples from the insect cell lysate expression system. Buffer containing ZnCl<sub>2</sub> (Figure A1.3B, lanes 5 and 6) was compared to the previous experiment’s buffer, which contained ZnSO<sub>4</sub> instead<sup>7</sup> (Figure A1.3B, lane 7). Full-length ovulin was visible in all samples (Figure A1.3B, lanes 5-8). I could see a very light band that may have been an ovulin processing product in the sample containing CG11864 null-mated female tracts and “active” CG11864 in the new buffer (Figure A1.3B, lane 6). However, this result was not reproducible in follow-up experiments using the same conditions with new samples.



**Figure A1.2: Ovulin and CG11864 expression with eukaryotic *in vitro* expression systems.** **A)** Ovulin Western blot. Lane 1: 1 pair male AG. Ovulin expressed in males is approximately 36kDa. Lane 2: ovulin expressed *in vitro* from the pExp1-DEST (N-terminal His tag) expression vector. Lane 3: ovulin expressed *in vitro* from the pExp2-DEST (no fusion tag included due to stop codon) expression vector. Both *in vitro* protein samples were obtained using the Qiagen Easy Xpress Insect Kit II. **B)** A Western blot, probed with anti-CG11864 antibody. Both the full-length and “active” (with pro-sequence deleted) forms of CG11864 were expressed using 2 *in vitro* kits: the Promega TNT-Coupled Wheat Germ Extract system and the Qiagen system used in **A**. Transcription/translation reactions were all done using pExp2-DEST with no fusion tag. Lane 1: 1 pair male AG. Lanes 2-5: *in vitro* Promega reactions using the expression vectors indicated. Lanes 6-8: *in vitro* Qiagen reactions using the expression vectors indicated. Ovulin reactions are included as negative controls. Arrows indicate bands most likely corresponding to full-length protein. Bands above those indicated by arrows are non-specific cross-reactive proteins.

<sup>7</sup> ZnSO<sub>4</sub> was used initially because ZnCl<sub>2</sub> was unavailable at the time. However, I decided to switch to ZnCl<sub>2</sub> because the first experiment was inconclusive, and the protocols I was working from used ZnCl<sub>2</sub>.



**Figure A1.3: Semi *in vitro* ovulin processing assays.** **A)** A Western blot, probed with anti-ovulin antibodies. Lane 1: 1 pair male AG. Lanes 2-4: reproductive tract extracts from wild-type females mated to wild-type males and dissected at times given after the start of mating (ASM). Lanes 5-6: Insect cell lysate kit-expressed CG11864 (full-length or “active”, as indicated), incubated with reproductive tract extracts from wild-type females mated to DGRP-517 males (CG11864 null). Lanes 7-8: Wheat germ extract kit-expressed CG11864 (full-length or “active”, as indicated), incubated with reproductive tract extracts from wild-type females mated to DGRP-517 males. Lane 9: Reproductive tract extracts from wild-type females mated to DGRP-517 males and dissected at 1 hour ASM. **B)** The experiment from **A**, repeated using only insect cell lysate kit-expressed CG11864 (see text). A “no plasmid” control (lane 8) is included to test for presence of enzymes that can process ovulin in the *in vitro* reaction mix. Arrow indicates position of a possible ovulin cleavage product in lane 6. Numbers to the left of the blots indicate approximate sizes in kDa.

### ***Baculovirus expression***

The cell-free expression kits, while convenient to use, do not produce large enough yields of protein to allow for purification or reliable detection of proteins on a Western blot. I therefore created baculovirus expression constructs for all proteins (ovulin, full-length CG11864, “active” CG11864, and seminease; see Table A1.1 for details) to express in Sf9 insect cells. I included the native signal sequences so that all proteins would be secreted into the culture media and post-translational modifications could be added through the secretory system. I used the pFastBac/CT-TOPO vector (Invitrogen) to clone all constructs. This vector includes a C-terminal 6x-His tag. I deleted stop codons from all inserts to allow addition of the His tag.

Viral stocks were obtained after transfection of Sf9 cells with the baculovirus construct, using the Cellfectin II reagent (Invitrogen). I re-infected cells with these initial viral stocks (P1) to obtain higher viral titers (P2-5). I tested protein expression on Western blots to determine the best titer to use. Early experiments used P5 viral stocks, in which seminease was not expressed. However, after my viral stocks were accidentally discarded, I repeated the transfection/infection cycle and determined that P3 was the best viral stock in all cases. Seminease was expressed with this new viral stock.

All proteins were successfully expressed and purified over a Ni-NTA column (Qiagen). The following protocol was used to pull down the His-tagged proteins from 30mL of culture



media at 72h post-infection:

1. Approximately 800 $\mu$ L Ni-NTA resin was rinsed with water to remove EtOH. Resin was then incubated in Wash Buffer (50mM Tris-HCL, pH 8.0, 300mM NaCl, 10mM imidazole, 10% glycerol) for 30 minutes.
2. Meanwhile, each 30mL suspension culture was centrifuged for 30 minutes at 4000 rpm. The pellet was discarded. A small aliquot of clarified media was saved as the **input** sample.
3. 30mL culture media was added to the resin and Wash Buffer from step 1. The media was incubated with the resin for 30-40 minutes with gentle shaking at room temperature.
4. The resin was pelleted with centrifugation for 10 minutes at 4000rpm. An aliquot of the **supernatant** was saved to test for proper binding of the proteins to the resin. The remaining supernatant was discarded after confirming pull-down of proteins.
5. The resin (with proteins bound) was transferred to an empty 10mL chromatography column. After allowing all excess media to run through, the column was washed with 10mL Wash Buffer 3 times.
6. Proteins were eluted with 3 1mL volumes of Elution Buffer (50mM Tris-HCL, pH 8.0, 300mM NaCl, 250mM imidazole, 10% glycerol).

I could detect reasonable levels of ovulin (Figure A1.4A), CG11864 (Figure A1.4B), “active” CG11864 (Figure A1.4C) and seminase (Figure A1.4D) on Western blots. However, I could not purify sufficient protein from 30mL of cell media to visualize on a Coomassie-stained gel.

CG11864 expressed in Sf9 cells migrated with slightly lower apparent molecular weight than did CG11864 in male accessory glands and both full-length and “active” CG11864 were the same apparent molecular weight (no AG sample in Figure A1.4B; other Westerns confirmed that both full-length and “active” forms of CG11864 had the same apparent molecular weight on a gel—data not shown). CG11864 may have been cleaved by other proteases present in the cell culture at its native cleavage site or a nearby cryptic cleavage site. The exact N-terminal sequence of the full-length CG11864 protein expressed in cell culture has not yet been determined (see below for discussion).

Ovulin expressed in Sf9 cells was glycosylated, but had the same apparent molecular weight as ovulin from male AG, even though it has a C-terminal 6x-His fusion tag (Figure A1.4A). To test for N-linked glycosylation of proteins, I treated samples of male AG extracts or Sf9-expressed ovulin with the enzyme PNGase F. PNGaseF removes N-linked glycosylation marks from proteins [222]. Ovulin from male AG samples shifted to a lower molecular weight after PNGase treatment, as expected (Figure A1.4E, lanes 1 and 2). Ovulin expressed in Sf9 cell culture also shifted to a lower molecular weight after treatment, becoming apparently the same size as deglycosylated ovulin from male AG (Figure A1.4E, lanes 3 and 4). These results indicate that either I was unable to separate the proteins in these samples enough to see a size difference (due to the C-terminal fusion tag) between native ovulin and my Sf9-expressed ovulin, or that some unknown modification of ovulin occurred during expression in cell culture.

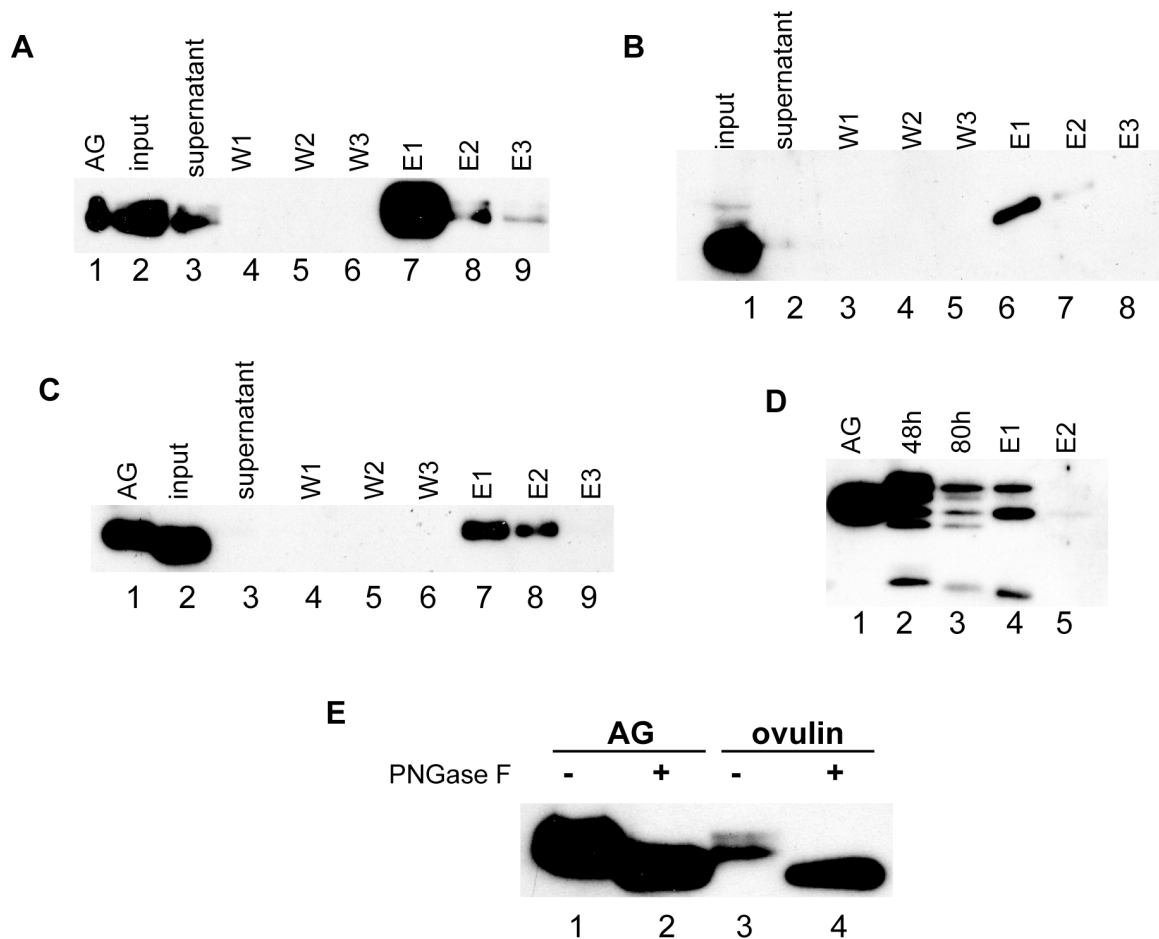
Both full-length and “active” CG11864 appear to be the same size on a gel, even though the “active” form should be approximately 3kDa smaller. This suggests that some modification of one or the other protein occurs after expression by Sf9 cells.

I performed semi *in vitro* processing assays by mixing protein extracts from female

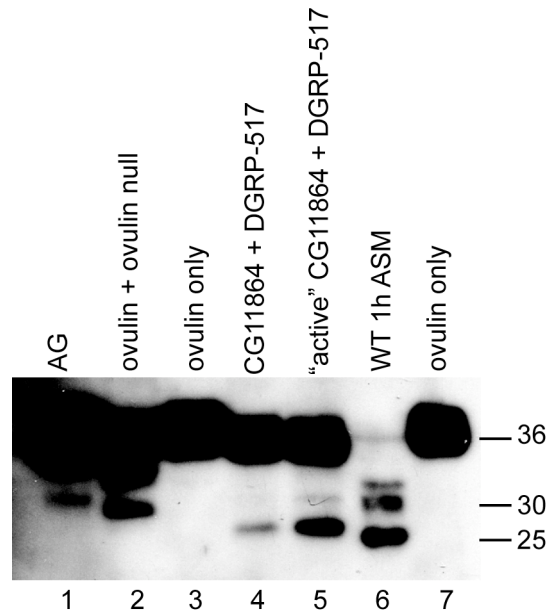
reproductive tracts after mating to mutant males (either ovulin null [105] or CG11864 null [Chapter 3]) with Sf9 cell-expressed proteins in their elution buffer. Female reproductive tracts were ground in 10 $\mu$ L Buffer A (described earlier) and mixed with an equal volume of protein eluate.

Sf9 cell-expressed ovulin was mixed with reproductive tract extracts of females that had been mated to ovulin null males. After 16 hours incubation at room temperature, lower molecular weight bands could be seen in this sample, along with full-length ovulin (Figure A1.5, lane 2). These band sizes roughly corresponded to those of the first and second ovulin cleavage products seen in females mated to wild-type males (Figure A1.5, lane 6). Sf9-expressed ovulin incubated alone overnight did not have any smaller molecular weight bands present (Figure A1.5A, lanes 3 and 7). This result indicated that ovulin expressed using the Baculovirus system could be cleaved by enzymes in mated females.

Sf9 cell-expressed CG11864 (both full-length and the “active” form) were mixed with reproductive tract extracts of females mated to CG11864 null males for 16 hours at room temperature. These samples both had bands that were slightly larger than the third cleavage product in wild-type mated females (Figure A1.5, lanes 4 and 5). There may have been slightly more of this potential cleavage product in the sample containing “active” CG11864, but because this result was not replicated, I could not be certain. This preliminary experiment suggested that, at least for a short time after purification, CG11864 expressed in insect cell culture was active. These experiments were all performed on the same day as protein purification (though not the same as the protein purification results shown in Figure A1.4). Follow-up experiments using these protein samples failed to replicate the results of this experiment.



**Figure A1.4: Proteins expressed in insect cell culture using the baculovirus expression system. A)** Ovulin Western blot. Lane 1: 1 pair male AG. Lane 2: Aliquot of input (clarified cell culture media at 72 hours post-infection with ovulin-expressing viral stock). Lane 3: Aliquot of supernatant after incubation of input with Ni-NTA resin. Lanes 4-6: Aliquots from three column washes. Lanes 7-9: Aliquots from three 1-mL elutions. **B)** CG11864 Western blot. Samples were collected from media of cells infected with full-length CG11864-expressing viral stock. Labels are the same as in A. **C)** CG11864 Western blot. Samples were collected from media of cells infected with viral stock expressing the “active” form of CG11864. Labels are the same as in A. **D)** Seminase Western blot. Lane 1: Protein from 1 pair male AG. Lane 2: Aliquot of cell media at 48 hours post-infection with seminase-expressing viral stock. Lane 3: Aliquot of cell media at 80 hours post-infection. Lane 4: Aliquot of first elution. Lane 5: Aliquot of second elution. **E)** Ovulin Western blot. Lanes 1-2: 1 pair male AG each. Lanes 3-4: 10 $\mu$ L of ovulin elution sample (as in A). +: treated with PNGase F.



**Figure A1.5: Semi *in vitro* ovulin processing assay with Sf9-expressed proteins.** Ovulin Western blot. Lane 1: 1 pair male AG. Lane 2: Sf9-expressed ovulin incubated at room temperature with reproductive tract extracts from females mated to ovulin null males. Lanes 3 and 7: Sf9-expressed ovulin, incubated at room temperature. Lanes 4-5: Sf9-expressed full-length or “active” CG11864 (as indicated above blot) incubated with reproductive tract extracts from females mated to DGRP-517 males (CG11864 null). Reproductive tract extracts from females mated to wild-type males and dissected at 1 hour ASM. All females used were wild-type. Numbers to the right of the blot indicate approximate sizes in kDa.

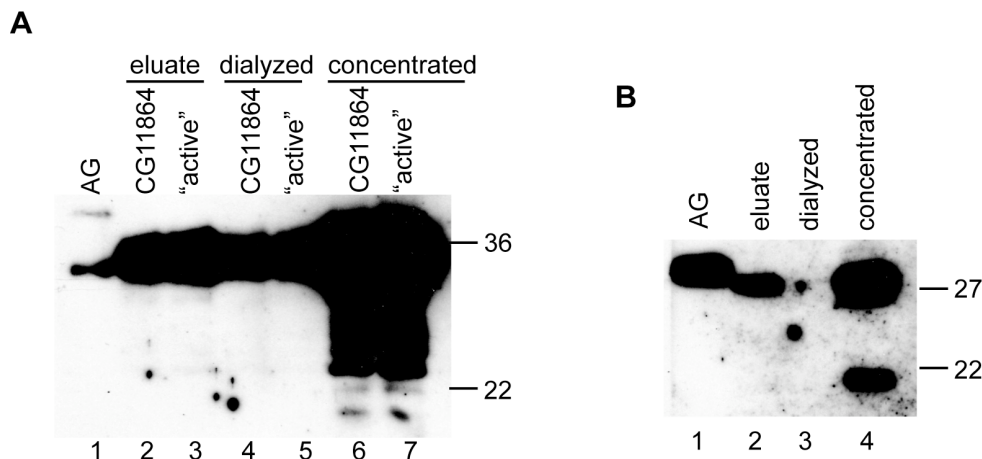
### ***Buffer exchange and protein concentration***

The buffer used to elute these proteins from the nickel column contained very high concentrations of both NaCl and imidazole, both of which may inhibit proteolysis. I therefore attempted to exchange buffer and purify CG11864 and seminase protein from separate 200mL cultures and subsequently concentrate the protein by traditional dialysis followed by centrifugation with 2mL filter columns (Millipore Centricon). Proteins were purified from the cell media over a Ni-NTA column prior to dialysis. I decided to focus my efforts on the proteases, as these were most likely to be the limiting factors for ovulin cleavage. Dialysis was performed in 1L buffer with 3mL eluted protein sample in dialysis tubing. Samples were kept at room temperature for approximately 6 hours, with buffer exchanged every 2 hours. Following this, the buffer was exchanged a final time and samples were kept at 4°C overnight. The buffer used for exchange was: 20mM Tris-HCl (pH 7.5), 150mM NaCl, 2.5mM CaCl<sub>2</sub>, 2μM ZnCl<sub>2</sub>, 0.004% Tween-20.

I was able to recover CG11864 protein after dialysis (Figure A1.6A, lanes 4 and 5), though dialysis lowered the concentration of seminase in the sample (Figure A1.6B, lane 3). All three proteins were concentrated by the filter tube (Figure A1.6A, lanes 6 and 7; Figure B, lane 4). However, I did not obtain sufficient protein to detect bands on a Coomassie-stained gel (not shown).

I mixed the concentrated enzyme samples with previously purified ovulin from Sf9 cells to test for ovulin cleavage. I did not see any evidence of ovulin cleavage in these samples (not

shown). Additionally, I attempted *in vitro* ovulin cleavage assays with purified proteins that were dialyzed directly in the centrifugation filter tubes. I did not see any ovulin processing products on a Western blot (not shown).



**Figure A1.6: Dialysis and concentration of Sf9-expressed proteins. A)** CG11864 Western blot. Lane 1: 1 pair male AG. Lanes 2-3: Protein eluate from Ni-NTA columns (see text). Baculovirus stocks used for Sf9 infection are shown above lanes. Lanes 4-5: Samples from lanes 2-3 after dialysis. Lanes 6-7: Samples from lanes 4-5 after concentration in spin columns. **B)** Seminase Western blot. Lane labels are the same as in A. Numbers to the right of blots indicate approximate sizes in kDa.

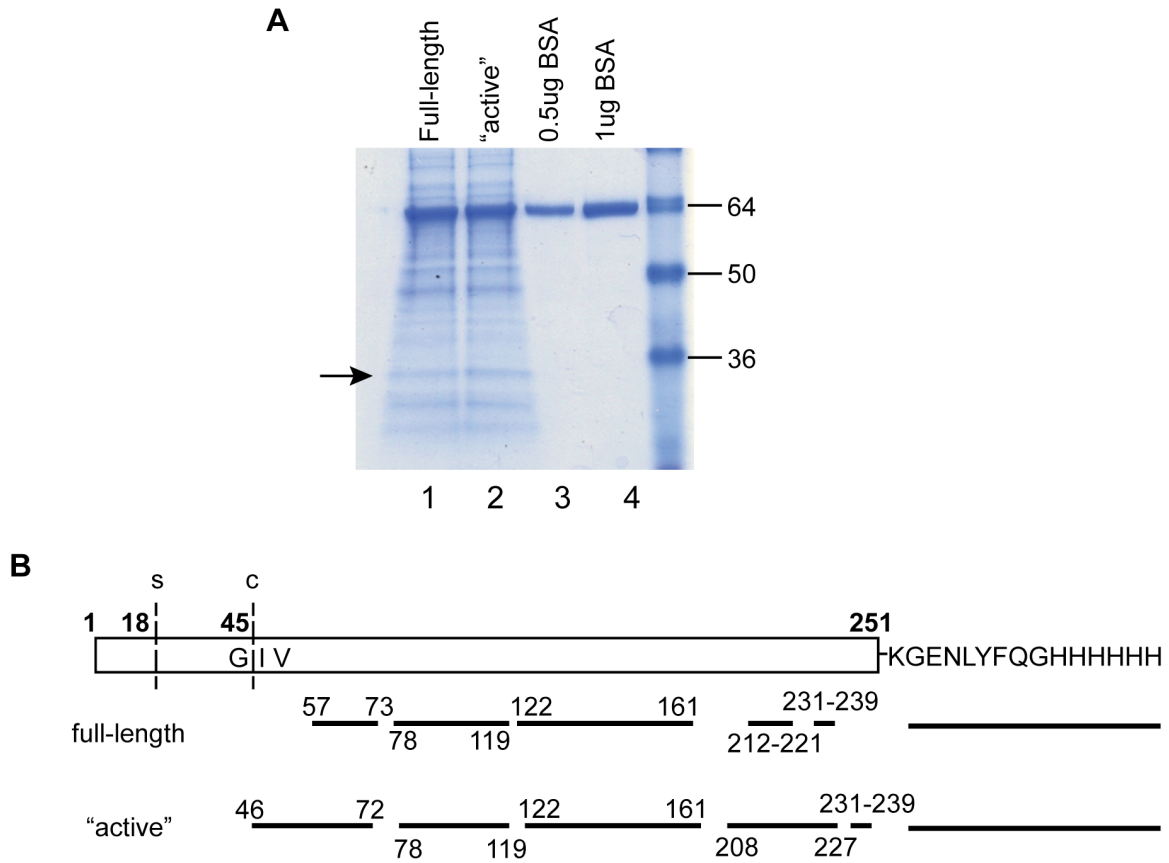
In a final experiment, I expressed CG11864 and “active” CG11864 in 30mL cultures and concentrated the media directly (i.e., I did not run the sample over a nickel column) to an approximately 1mL volume using Centricon 2mL filter tubes (Millipore). I was able to see a band of the correct size on a Coomassie-stained gel, though it was still very weak (Figure A1.7A). Bands were excised and submitted for N-terminal mapping to the Proteomics and Mass Spectrometry core facility. Secreted “active” CG11864 matched its predicted N-terminal sequence (Figure A1.7B). However, there was not enough protein from the N-terminal region of the full-length CG11864 sample to accurately map its N-terminus (Figure A1.7B).

## CONCLUSIONS AND FUTURE DIRECTIONS

These experiments demonstrate that ovulin, CG11864, and seminase can be expressed *in vitro* using various systems. Expression of “active” CG11864 using insect cell-free lysates and subsequent mixing with mated female reproductive tracts that do not contain CG11864 yielded some promising results as ovulin appeared that it may have been partially processed in at least two samples (Figure A1.3). However, the low protein expression in this system is likely responsible for the low reproducibility.

Expression of proteins in Sf9 insect cells using the baculovirus expression system yields enough protein for Western blots, but this may not be enough for *in vitro* assays. Enzymes expressed in this system may not be as efficient as those expressed in the male accessory gland, so larger amounts of protein may be required for ovulin cleavage. I was only able to perform a

limited number of expression and purification experiments. Future experimenters may be able to optimize the protocols for larger and more concentrated purified protein yields. In addition, N-terminal mapping of the “active” CG11864 protein should be repeated to determine if the predicted protein is expressed, or if secondary cleavage of the protein occurs after secretion into the cell media.



**Figure A1.7: N-terminal mapping of Sf9-expressed CG11864 and “active” CG11864.** **A)** Coomassie-stained acrylamide gel. Lane 1: Approximately 40μL of full-length CG11864, from concentrated of raw cell culture media. Lane 2: The same as Lane 1, except with the “active” CG11864. Lane 3: 0.5μg BSA. Lane 4: 1μg BSA. Arrow indicates band containing CG11864 (as determined using a Western blot with identical samples on the same gel—not shown). Numbers to the right of gel indicate approximate sizes in kDa. **B)** N-terminal mapping performed with bands cut out of a Coomassie-stained gel similar, though not the same gel, as that shown in A. The diagram illustrates the protein sequence of CG11864 expressed in cell culture. This protein should be identical to native CG11864, with the addition of the amino acids shown on the right (C-terminus). “s” is the location of the signal sequence cleavage site. “c” is the location of the predicted pro-peptide cleavage site (and the predicted N-terminus of the “active” recombinant protein). Peptides mapping to the CG11864 protein sequence are shown below the diagram for each protein sample submitted. Numbers correspond to CG11864 numbering used in the diagram and indicate termini of each peptide.

**Table A1.1: List of constructs mentioned in Appendix 1**

Plasmid name	Plasmid description	Portion of inserted gene included (based on CDS numbering, in bp)	Fusion tag	Antibiotic resistance
Acp26Aa pExp1-DEST	full-length ovulin <i>in vitro</i> expression vector (no signal sequence)	55-stop (795)	N-term 6xHis	amp
CG11864 no sig pExp1-DEST	full-length CG11864 <i>in vitro</i> expression vector (no signal sequence)	55-stop (756)	N-term 6xHis	amp
CG11864 active pExp1-DEST	pro-delete CG11864 <i>in vitro</i> expression vector (no signal sequence)	136-stop (756)	N-term 6xHis	amp
CG10586 nosig pExp1-DEST	full-length seminase <i>in vitro</i> expression vector (no signal sequence)	58-stop (828)	N-term 6xHis	amp
Acp26Aa pExp2-DEST	full-length ovulin <i>in vitro</i> expression vector	full CDS (including stop)	C-term 6xHis (not expressed)	amp
CG11864 pExp2-DEST	full-length CG11864 <i>in vitro</i> expression vector	full CDS (including stop)	C-term 6xHis (not expressed)	amp
CG11864 prodel pExp2-DEST	pro-delete CG11864 <i>in vitro</i> expression vector	1-54, 136-756 (including stop)	C-term 6xHis (not expressed)	amp
CG10586 pExp2-DEST	full-length seminase <i>in vitro</i> expression vector	full CDS		amp
pFastBac/CT-TOPO Acp26Aa	donor plasmid for bacmid generation with full-length ovulin insertion	full CDS, no stop codon	C-term 6xHis	amp
pFastBac/CT-TOPO CG11864	donor plasmid for bacmid generation with full-length CG11864 insertion	full CDS, no stop codon	C-term 6xHis	amp
pFastBac/CT-TOPO CG11864 prodel	donor plasmid for bacmid generation with CG11864 pro-delete insertion	1-54, 136-753 (no stop codon)	C-term 6xHis	amp
pFastBac/CT-TOPO CG10586	donor plasmid for bacmid generation with full-length seminase insertion	full CDS, no stop codon	C-term 6xHis	amp
Acp26Aa bacmid	ovulin bacmid glycerol stock in DH10Bac E. coli (use bacmid for generating baculovirus stock)	full CDS, no stop codon	C-term 6xHis	kan, gentamicin, tet
CG11864 bacmid	CG11864 bacmid glycerol stock in DH10Bac E. coli (use bacmid for generating baculovirus stock)	full CDS, no stop codon	C-term 6xHis	kan, gentamicin, tet
CG11864 prodel bacmid	CG11864 pro-delete bacmid glycerol stock in DH10Bac E. coli (use bacmid for generating baculovirus stock)	full CDS, no stop codon	C-term 6xHis	kan, gentamicin, tet
CG10586 bacmid	seminase bacmid glycerol stock in DH10Bac E. coli (use bacmid for generating baculovirus stock)	full CDS, no stop codon	C-term 6xHis	kan, gentamicin, tet

APPENDIX 2  
MOLECULAR SOCIAL INTERACTIONS:  
*DROSOPHILA MELANOGASTER* SEMINAL FLUID PROTEINS AS A CASE STUDY<sup>8</sup>

## INTRODUCTION

While studies of social behavior generally focus on observable interactions between individuals, additional “hidden” social interactions occur on the molecular level. These molecular interactions can be considered social in two ways. First, observable social interactions are influenced by molecular interactions [270]. Second, molecules from different individuals can interact in what we call here “molecular social interactions”. The molecular biology of social behavior has thus far been focused primarily on the former: molecular interactions within an animal that either induce or result from social interactions. This approach has successfully identified molecular interactors in rodent and avian affiliative behavior (reviewed in [271, 272]), nematode feeding behavior (reviewed in [273]), eusocial behavior [274], and *Drosophila* courtship (reviewed in [275, 276]). However, a complete molecular understanding of social behavior necessitates an understanding not just of how molecules interact within a social animal, but also how “social molecules” interact among animals.

Here, we present a case study of such “molecular social interactions” that involves *D. melanogaster* seminal fluid proteins (Sfps) that are produced in the male reproductive tract, and transferred to the female along with sperm during mating. In the case of *Drosophila melanogaster* Sfps, the molecular social interactions are extensive, as gene products in seminal fluid induce short- and long-term changes in females’ behavior, physiology, and gene expression, and these changes require interactions of Sfps with female-derived molecules and physiology (e.g., muscle, circulatory, and neural systems). Thus, the male- and female-derived molecules are involved in an inherently social interaction—that is an interaction between two individuals of the same species. Molecular social interactions affect the outcome of individual matings and occur directly between males and between males and females, and indirectly between multiple males that have mated with a given female. As we will discuss, molecular social interactions both shape, and are shaped by, observable behavioral interactions between conspecifics to affect lifetime reproductive success.

Following mating, female *D. melanogaster* display a number of behavioral and physiological changes that impact both male and female reproductive success. For example, after mating, females increase their rates of oogenesis, ovulation, egg laying, and food intake (reviewed in [226, 277, 278]). Sperm from the male are stored in specialized sperm storage organs, and this process may be facilitated by changes in uterine shape beginning at the onset of mating [135, 206]. For several days, mated females are less likely to accept suitors, actively fleeing or kicking any persistent male [279, 280]. Within hours after mating, the female increases expression of several known anti-microbial peptide genes [185, 186, 281-283], yet the realized immune response that protects the female from infections is reduced [284]. The lifespan of *Drosophila* females is also reduced by mating [172, 173, 285-287].

These changes in behavior, physiology, and gene expression may be brought about by the

---

<sup>8</sup> A version of this chapter appeared in Sirot, L.K., LaFlamme, B.A., *et al.* (2009) *Advances in Genetics* Vol. 68, pp. 23-56. I have included in this appendix only the text for which I had primary responsibility. This excerpt is reprinted here with permission.



behavioral act of mating, by the transfer of sperm, or by other contents of the seminal fluid. Since males that do not produce sperm still elicit post-mating responses in their partners (albeit, weaker and more short-term) [202, 266, 288, 289], non-sperm components of the seminal fluid must be involved in the induction of these responses. In fact, males that transfer sperm but do not transfer Sfps produced in their accessory glands fail to elicit most post-mating responses in females [202, 266]. It is known that the ejaculatory duct and ejaculatory bulb also produce secreted proteins that constitute part of seminal fluid, and that some of these proteins are necessary for post-mating responses [290-293].

Sfps comprise an elaborate intraspecific signaling system. Of the more than 180 predicted extracellular proteins present in the reproductive secretory glands of male *D. melanogaster*, over 100 have been confirmed to be transferred to the female along with sperm (reviewed in [200, 294]; see also [56-59, 62]). Many of the transferred proteins fall into conserved protein classes found in the seminal fluid of most animals studied to date and include proteases, protease inhibitors, acid lipases, cysteine rich secretory proteins (CRISPs), and lectins [31, 200]. Other, less-expected, classes of Sfps such as odorant binding proteins suggest a possible role for small molecules in inducing female post-mating responses [57]. Odorant binding proteins are known to shuttle pheromones or other small molecules to odorant receptors in the olfactory system (reviewed in [295]). Presence of predicted odorant-binding proteins in the seminal fluids suggests that they may play a similar shuttling role for molecules once within the female reproductive tract. The wide variety of protein classes present in the seminal fluid suggests that Sfps take part in a complex series of interactions within the mated female and do not just fulfill a single simple role.

Upon transfer to females, Sfps target to specific tissues which are likely to relate to their function within the mated female [20, 76, 78, 189, 205, 291]. For example, proteins associated with sperm storage and retention have been detected in the female sperm storage organs, and ovulin, which stimulates ovulation, targets to the base of the ovaries [20, 78]. Several Sfps, including ovulin, have also been detected in the circulatory system of mated females from where they can gain access to the brain and/or endocrine systems [76, 78, 291, 296] and thus, potentially, affect female behavior. Further studies of the targets of Sfps may help to uncover their functions in the mated female.

*Drosophila melanogaster* Sfps provide an excellent model system in which to investigate molecular social interactions, due to the powerful tools available in this species. Mutant or transgenic males in which Sfps are increased, decreased, or eliminated can be used to dissect the effect(s) of particular Sfps on female post-mating responses [e.g., 6, 8, 67, 80, 81, 105, 129, 203, 204, 290, 297]. A large collection of freely available genomic databases (e.g., FlyBase; Fly atlas [62]) facilitate rapid progress as well. These techniques and tools, along with studies associating allelic variation in Sfps with variation in their effects, have led to a greater understanding of the molecular social interactions taking place between all of the players involved in *Drosophila* mating (reviewed in [226]). Furthermore, studies of *D. melanogaster* Sfps are likely to provide insights into the molecular social interactions of other species given that Sfps impact female post-mating responses across a wide taxonomic range (reviewed in [1, 24]).

While over 180 known or putative *D. melanogaster* Sfps have been identified, only one female receptor to an Sfp is known: the sex peptide receptor, a G protein-coupled receptor expressed in the female reproductive tract and nervous system [298]. However, we expect that many Sfps interact with female-derived proteins. Some female-derived proteins that play a role in female post-mating behavior and physiology have been identified and will be discussed in this

review, but their interactions with Sfps remain speculative at this time.

Several approaches have been used to identify genes in females whose products mediate response to, are regulated by, or otherwise interact with, Sfps. Proteins produced in the female sperm storage organs have been identified and have the potential to interact with Sfps [187, 190, 299, 300]. Microarray data from whole flies, heads, or reproductive tract tissues have shown that different aspects of mating, including Sfps, cause a transcriptional response in the female after mating [185-188, 281-283, 301, 302], though it is not likely that most post-mating responses are due to mating-induced transcription. Transcriptional changes of the largest magnitude are seen by about 6 to 8 hours after mating, a time by which most Sfps are no longer detectable in the female. Therefore, Sfps may set into motion the transcriptional modification of the female, but the genes regulated by these modifications are less likely to encode Sfp-interacting proteins than the genes expressed by the female prior to mating. Nevertheless, these mating-regulated genes likely are players in the next steps of the molecular social interactions. To fully understand the molecular social interactions in which Sfps are involved, we must identify female interactors, their functions, and how they have co-evolved with their male-derived partners.

## BEHIND THE SCENES: EVOLUTIONARY DYNAMICS

Behind the scenes of the main mating arena, a number of important factors have been setting the stage. Some of these are developmental genes that determine the sex of the fly, such as *Sxl* (reviewed in [303-305]) or its downstream effectors, such as *dsf* and *fru*, that are needed for sex-specific behavior and neural development (reviewed in [276, 305-307]). Other genes are required for proper development of the genital tract and fertility; examples include: *HR39* [299] or *lz* [308] in females, and a late function of *prd* [215] in males. Though the actions of these genes are thus ultimately important in mating interactions of *Drosophila*, they have a limited role in the types of molecular social interactions we have discussed and will not be discussed further here. On the other hand, evolutionary processes shape all genes in the genome and have a specific role in affecting mating behavior and related social interactions in *Drosophila*. These are discussed further, below.

Molecular social interactions, such as those discussed earlier, influence not only the individual interacting flies but also the evolution of the interacting molecules. Over time, evolutionary responses to the interactions between Sfps and female molecules may leave signatures on the genes' sequences themselves (some examples in *Drosophila* include: [29, 30, 55, 66, 75, 193, 309, 310]; for reviews, see [27, 311]. Genes encoding *Drosophila* Sfps, like reproduction-related genes across a wide range of taxa, are more likely to show evidence of positive selection than are most other groups of genes, with the notable exception of immunity-related genes [30, 55, 66, 75, 193, 309, 310, 312, 313]; for reviews, see [27, 29, 311]. It has been suggested that rapid evolution of these genes may result from sexual selection (e.g., cryptic female choice, sperm competition) and/or sexual conflict (reviewed in [27]).

An example of the interplay between mechanisms of action and evolutionary dynamics is ovulin. Earlier, we presented ovulin processing as an example of a highly coordinated process between the sexes that begins in the male and ends in the female. Thus, ovulin's action to increase ovulation soon after mating would appear to benefit both sexes, but further consideration raises some questions. Is ovulin a mechanism by which the male controls the female's physiology, or does the female take advantage of the male's ovulin contribution for her own benefit? Has ovulin's evolution been driven by sexual conflict? Or, is ovulin's function

truly beneficial to both sexes? Though we currently cannot answer many of the important questions about ovulin, such as why it is processed or how processing of this protein affects its function, we may be able to gain some clues from the evolutionary history of the protein.

Consistent with its important function in reproduction, ovulin contains short regions that are highly conserved between *Drosophila* species [314, 315], with its C-terminus in particular containing several conserved motifs. Genetic and biochemical studies indicate that these motifs mediate self-association of ovulin molecules, evidenced by observations of ovulin multimers *in vivo* in *D. melanogaster*, as well as *in vitro* between ovulin molecules from different species that contain these motifs [314]. These self-association domains are within a portion of ovulin that has ovulation-inducing activity, although it is as yet unknown whether self-association is needed for ovulin's function.

Interspersed around the highly-conserved backbone of ovulin are rapidly-evolving regions. In fact, ovulin is one of the most rapidly-evolving proteins in the *Drosophila* genome [211, 316-318]. Amino acid divergence between ovulin from *D. melanogaster* and its close relative, *D. simulans*, is about 15%, compared to a 1-2% average for all other genes [319-321]. There is evidence that some of ovulin's amino acid divergence results from positive selection, but the selective forces have not been identified [211, 316-318, 322]. The ovulin-type pattern of highly divergent protein regions on an otherwise conserved backbone may be the result of sexual conflict, resulting from sexual selection, acting on proteins required for essential biological processes.

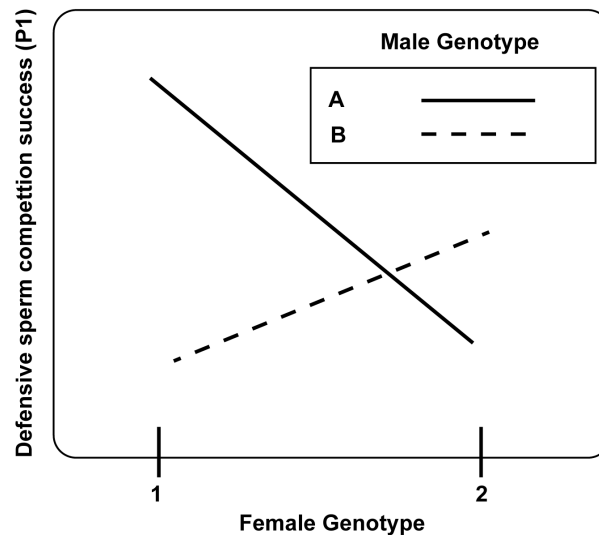
Given the importance of Sfp genes to reproductive success, positive selection should drive the most advantageous alleles to fixation. We expect any variation in genes under positive selection to be transient. However, in contrast to this prediction, nucleotide variation is maintained in some Sfps, like ovulin [316, 318], and in Sfp-mediated traits, like sperm competition [277, 323, 324]. Why is so much polymorphism maintained?

The answer may be that there is no single "most advantageous" allele for a particular Sfp. What is "best" may depend on the genotypes of the individuals involved in the interaction. Using sperm competition as an example, within the mated female a male's sperm must compete for fertilizations with sperm from other males with whom the female had mated or will mate. In the case of three different genotypes (A, B, and C), Male A may outcompete male B, and male B may outcompete male C, but this does not mean that male A will outcompete male C ( $A > B$ ,  $B > C$ , but  $C > A$ ), when female genotype is held constant. This property of sperm competition, known as nontransitivity [324], tells us that the outcome of a sperm competition interaction depends on the particular Sfp alleles of the competing males. Antagonistic pleiotropy may also lead to maintenance of polymorphism among Sfp genes [79, 325]. For example, a polymorphism in the Sfp CG17331 is associated both with offensive sperm competitive ability and female refractoriness. When males carry the "C" allele, they are better able to prevent their mates from re-mating, but are poorer at displacing a previous mate's sperm [325]. The dependence of female post-mating behavior on the specific genotypic combinations of her multiple mates makes it difficult for one Sfp gene allele to "win".

The female's genotype, along with the interaction between male and female genotypes, also determines the outcome of sperm competition (Figure A2.1) [326, 327]. Complicating matters further, there is no genetic correlation between male-derived and female-derived effects on sperm competition [328]. When the genotypes of competing males are held constant, a change in female genotype can determine which male is the "winner". The complex interactions between the genotypes of competing males and between male and female genotypes would make it nearly

impossible for a particular male genotype to persevere in sperm competition. In fact, artificial selection even fails in the face of this complexity [329].

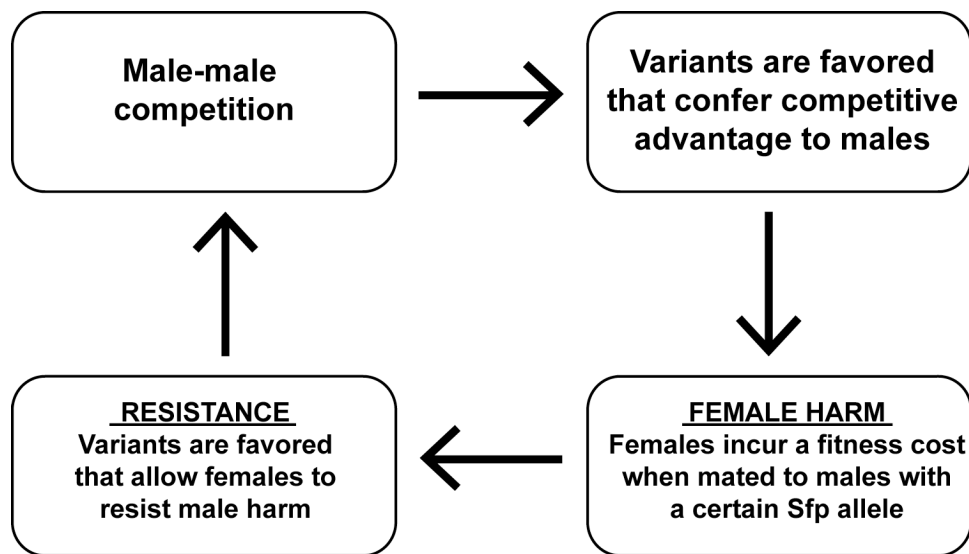
Males must compete against each other for fertilization success and against the female's sometimes conflicting interests. For example, one might expect males to benefit from inhibiting females from remating, whereas females may benefit from receiving the sperm of multiple males. It is also likely, given the differing magnitudes of investment in progeny production between males and females, that they have different optima for the level of post-mating effects (such as ovulation rate) [330]. Conflict between the sexes due to these different optima has been interpreted to suggest a kind of evolutionary “arms race” (Figure A2.2). First, male-male competition selects for alleles conferring higher male competitive reproductive success. These alleles may lower a female's lifetime progeny production for unknown reasons, resulting in selection for females to counter-adapt to overcome harmful male alleles [330-333]. At any given time in a population, this scenario of adaptation and counter-adaptation would occur simultaneously for many traits, leading to high levels of allelic polymorphism at the loci involved. Consistent with this prediction, we see that maintenance of variation in male-derived proteins depends on female genotype (reviewed in [277]). Furthermore, variation in Sfp-mediated female post-mating traits, such as female sperm usage patterns (determined by sperm competition experiments), is also dependent on female genotype [301, 309, 327, 334].



**Figure A2.1: Male and female genotypes affect outcome of sperm competition experiments.**

The genotypes of the particular male and female involved in a mating affect the defensive component of sperm defense, P1. P1 measures the proportion of offspring sired by the first male after the female has remated. In this hypothetical example, male A has a much higher P1 than male B when either of them mates with a female of genotype 1. However, when these same males mate a female of genotype 2, male B now has the advantage. Here, the second male to mate in each competition experiment does not vary. These interactions become even more complex when the genotype of the second male to mate is taken into consideration.

Experimental evolution studies support the hypothesis that evolution of Sfps is driven, in part, by sexual selection. When monogamy is enforced on *D. melanogaster* for multiple generations, males become less harmful and females become less resistant to harm than polygamous controls [335]. In *D. pseudobscura* populations selected under monogamy, normal levels of promiscuity, or elevated levels of promiscuity, males differed in their ability to prevent remating of their mates [336]. Counterintuitively, *D. pseudobscura* males selected under monogamy induced greater refractory periods in females, suggesting that the males selected under promiscuous female conditions invested more in pre-copulatory competition than in post-copulatory competition. In another experiment, when female *D. melanogaster* were prevented from co-evolving with a male population that was allowed to adapt to the static female genome, these males were able to induce higher remating rates in females and were better at the “defense” component of sperm competition [337]. Though we do not know for certain whether any of these results are due to changes in Sfps (either at the sequence or expression levels), the study by Wigby *et al.* [338] suggests that Sfps are involved in these types of adaptations because, in that study, changes in levels of transferred Sfps correlated with changes in post-copulatory traits in females (see also [339]).



**Figure A2.2: Sexual conflict may maintain polymorphisms in the population.**

Competition among males for mates/fertilizations drives selection for new alleles that confer a competitive advantage to males for some reproductive trait. For example, males with this allele may induce a higher rate of ovulation in females by way of a particular Sfp. This higher ovulation rate may be harmful to females due to higher energy costs, driving selection for alleles that allow females to resist the effects of the harmful male allele. This cycle continues, with males developing new competitive strategies, each with a potential female cost associated with them. As a result of this process occurring simultaneously at many loci, polymorphisms are maintained in the population. Adapted from Arnqvist & Rowe 2005.

## FEMALE CONTRIBUTIONS TO MOLECULAR SOCIAL INTERACTIONS

Since many Sfps are rapidly evolving, we might expect parallel patterns of evolution for the many female proteins (yet to be identified, with the exception of SPR) with which they interact. In line with this prediction, there are rapidly-evolving genes expressed in the female reproductive tract [187, 300, 340], although this group as a class is not enriched for rapidly-evolving genes [30]. However, molecular evidence for the type of “arms race” between the sexes, mediated by Sfp-female interactor pairs, is lacking. The only female interactor of an Sfp that has been identified, the G-coupled protein receptor SPR, has not undergone rapid evolution; in fact, it is strikingly conserved across evolutionarily distant lineages [298]. Studies that have identified rapidly-evolving female reproductive tract proteins found that they are enriched for serine proteases, not receptors as might have been predicted [312, 341], though several rapidly evolving female genes also have unknown functions. Some serine proteases expressed in female reproductive tracts are differentially expressed in response to mating [187, 190], suggesting a possible direct or indirect interaction with Sfps.

Much work needs to be done to determine whether SPR is the norm in *Drosophila*, or an exception. Examples can be found from many other species where, unlike SP and SPR, variation in a male or female molecule seems to drive rapid evolution at its partner from the opposite sex. For instance, the abalone sperm lysin protein has undergone rapid diversification, which is thought to be a response to sequence changes in its receptor on the egg [312, 341, 342]. Before we can test the prediction that changes in male reproductive proteins cause changes to female molecules (or vice versa) and downstream behaviors, more male/female molecule pairs must be identified. Transcriptional profiling studies of females after mating [185, 186, 188, 283], in conjunction with evolutionary studies, will likely lead to identification of further promising candidates.

## DISCUSSION AND FUTURE DIRECTIONS

We now know a great deal about how Sfps affect social interactions related to mating in *D. melanogaster*. However, the current body of research on Sfps is only the tip of the iceberg when it comes to understanding how the social interactions themselves affect Sfp usage and evolution. Future experiments can explore new social contexts in *D. melanogaster* for their effect on the overall amount of Sfps transferred, the effect on specific Sfps, and the ultimate reproductive consequences for each individual involved. Experimental evolution studies may test for changes in Sfp production or mating-induced transcriptional responses in response to altered population dynamics.

One of the greatest challenges to elucidating the relationship between organismal level interactions and individual Sfps (and their female counterparts) is the difficulty in performing experiments under “natural” conditions. Advantageous reproductive traits no doubt depend on environmental conditions. For example, the larval environment [343] and adult male nutrition [344] have been shown to affect post-mating traits in *D. melanogaster*. In the lab, there are no predators and flies do not experience the dangers they might in nature, such as depletion of food supply, extreme temperatures, or desiccation. Simulating as wide an array of conditions as possible will provide valuable information on which social interactions are really important in nature, and how these interactions affect Sfp dynamics and post-mating traits.

Several studies have considered at least one of the important conditions experienced in nature: genetic variation among individuals. While most mechanistic studies of Sfps have been carried out using standard lab strains, variation in sperm competition and other post-mating traits have been linked to variation in male and female genotypes [79, 301, 323, 327, 345, 346]. Currently, the genomes of nearly 200 inbred strains derived from wild-caught *D. melanogaster* are being sequenced [347]. This unprecedented genomic tool will provide a means for identifying the genetic basis of any number of complex traits, including behaviors affected by Sfps. Future experiments will be able to tease apart the effects of different genotypic combinations on social interactions such as receptivity to mating and sperm competition. From here, we can finally start to understand how females, not just males, act to control their reproductive success and to affect the reproductive success of their mates. Only then will we be able to comprehend how all levels of mating interactions come together to affect the complex molecular social interactions between *Drosophila* individuals.

### APPENDIX 3

## INVESTIGATING THE ROLE OF MATING IN EXPRESSION OF *DROSOPHILA* OOGENESIS-RELATED GENES

### OVERVIEW

In the fruit fly *Drosophila melanogaster*, male-derived Accessory gland proteins (Acps) cause a number of physiological and behavioral changes to females following mating. One of the most striking post-mating changes in females is that they greatly increase their rates of oogenesis and oviposition [reviewed in 1-3, 24]. Although Acps responsible for this change are known, such as the Acp ovulin and sex peptide [20, 105, 203, 204], the molecular basis of their action is not well understood.

In sexually mature virgin females, oogenesis proceeds at a very low rate, with some oocytes fully developed but not ovulated, and most of the remainder blocked at stage 9 of oogenesis [348]. Mating releases a block of unknown mechanism that allows the younger oocytes to proceed to stage 10; from there they can easily complete oogenesis [348]. The increase in stage 10-12 oocytes in the ovary that begins at about 10 hours after mating is dependent on receipt of Acps [105, 215].

Microarray studies of whole mated females identified genes whose expression levels were modified by mating [185, 187, 188] and by Acps in particular [185, 188]. Among these were genes expressed during oogenesis; oogenesis genes account for the majority of genes upregulated more than two-fold at 12 and 24 hours after mating [185]. Some Acps, including ovulin and Acp62F, localize to the ovary or the surface of laid eggs [78], where they may function in regulating ovulation or oogenesis.

Several chorion-related genes (structural proteins of the *Drosophila* eggshell) are differentially expressed in female *Drosophila* after mating to either wild-type males or males deficient in a single accessory gland protein (Acp) [188]. The eggshells of oviparous insects such as *Drosophila* are exposed to the pressures of both natural selection and sexual selection. They serve as the physical barrier between the developing embryo and the environment and are also the location of gamete-gamete interactions between males and females. Many genes that encode structural elements of the *Drosophila* chorion have been shown to be rapidly evolving, relative to vitelline membrane protein genes, likely due to these pressures [349].

I investigated how mating affects expression levels of chorion protein genes and other oogenesis-related genes, such as ecdysone-pathway genes responsible for chorion protein gene expression [350], at specific stages of oogenesis to start to dissect how mating may lift the stage-9 block to oogenesis. I measured transcript levels of 10 genes in early and late oogenesis in both mated and virgin females, and 6 genes in females mated to Acp62F null males versus controls. Overall, gene expression levels accurately reflected the oocytes stages in the samples. Neither mating nor loss of Acp62F caused a significant difference in expression levels in this study, likely reflecting the small sample size. However, some small differences were detected. These data may be useful for future work studying the expression of these genes after mating.

### METHODS

#### *Effect of mating on oogenesis gene expression*

Flies were reared on standard yeast/glucose media. Vials used for aging virgin females



were supplemented with live yeast. I placed 3-5 day old virgin Canton-S (CS) females individually in vials with food. A single 3-5 day old virgin CS male was placed into each of half of the vials, allowing half the virgin females the opportunity to mate. Mated females were frozen 10 hours after mating, in parallel with virgin females. Females were kept at -80°C until dissection.

Flies were separated into two experiments. In the first experiment, I separated stage 1-12 oocytes (immature oocytes) from stage 13-14 oocytes (mature oocytes). I pooled stages 13 and 14 due to the difficulty in separating these two stages. This experiment was designed to confirm some of the results from the microarray studies by McGraw *et al.* [185, 188] in which late chorion proteins (e.g., Cp15 and Cp18) were up-regulated after mating. Late chorion proteins are expressed in stage 14 oocytes [351].

In the second experiment, stage 1-9 oocytes were separated from stage 10 oocytes. The purpose of this experiment was to determine whether any genes were specifically up- or down-regulated immediately past stage 9 block in mated females compared to virgin females.

I extracted total RNA from two biological replicates of each sample and made cDNA from 500ng RNA each using oligo dT primers. I performed quantitative RT-PCR (qRT-PCR) for 14 oogenesis-related genes and 1 housekeeping gene (*gpdh*). Table A3.1 shows the primer sequences used for each gene. Table A3.2 shows information pertaining to each gene, such as known function and expression changes in response to mating.

I measured standard curves for each primer pair. I included three technical replicates at each of four dilution points. Standard curves for each of the PCR products are shown in Figure A3.1. I calculated an efficiency score for each primer pair, using the methods described in the Qiagen brochure “Critical Factors for Successful Real-Time PCR (link)”. Efficiency was calculated as  $10^{-1/\text{slope}} - 1$ . Ideally, this score would be 1. I did not use genes with an efficiency less than 0.8 or greater than 1.1 for further analysis.

I normalized samples to the housekeeping gene *gpdh*. First, the Ct value (“critical threshold”, the PCR cycle at which amplification is detected and the reaction is in the log-linear phase) was fitted to the standard curve of the corresponding gene. Second, the fitted value was divided by the fitted value of *gpdh* in the same biological sample.

I used primer pairs for the following genes (those with satisfactory standard curves): *Cp15*, *Cp16*, *Cp18*, *Cp19*, *Cp38*, *E74*, *E75*, *EcR*, *femcoat*, *Pxt*. Slopes and efficiency scores corresponding to the standard curves of each primer pair are shown in Table A3.3.

### ***Effect of Acp62F on oogenesis gene expression***

Acp62F is an Sfp that localizes to the eggshell in eggs laid within a day after mating [78], so I chose to look at the effect of this protein on the expression of genes whose products are part of the eggshell. I performed matings as in the experiments described above, except that all females were mated either to Acp62F null males or heterozygous controls [80] and frozen at either 2 hours or 24 hours after the start of mating. I extracted total RNA from whole ovaries and qRT-PCR performed as above. I used the housekeeping gene *Rpl32* instead of *gpdh*.

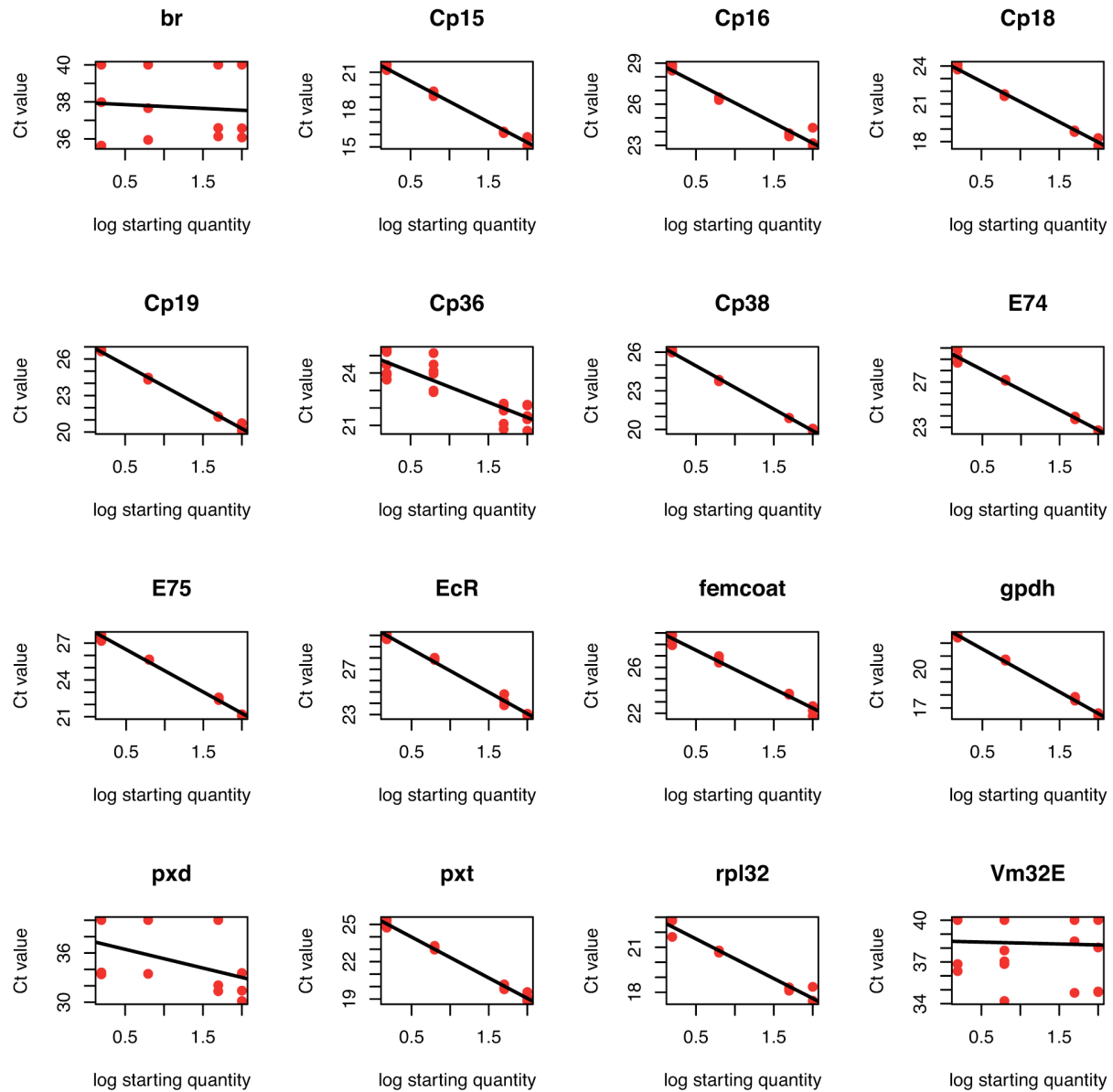
For this experiment, I only examined chorion protein genes (Acp62F localizes to the eggshell where these genes’ products are localized): *Cp15*, *Cp16*, *Cp18*, *Cp19*, *Cp38*, *femcoat*.

**Table A3.1: Primers used for qRT-PCR of oogenesis genes.** Primer sequences for each gene tested using qRT-PCR are shown here.

Gene	Primer 1	Primer 2
<i>BR-C</i>	CTGCATCAGTCGCTCAGTTC	CCAACGATCGGAGAGAGAAG
<i>Cp15</i>	TGAATCCCGGAACCTACAAG	TCAGAGTTTCGCCATGACAG
<i>Cp16</i>	TGACTGGAATTCCCTCAACC	TCTTCAGGTACGCATTCTCG
<i>Cp18</i>	GAGCTCCTCTGGTCTGGAAG	GTAGCTGGGCCTCTTGTAGC
<i>Cp19</i>	TCGCTACTCTGGCAGTCATC	GCTCATAACGCTGCTGACC
<i>Cp36</i>	CGAGGAGTACGTGAATGTGG	GCTTCTGGTAGTTGGGATCG
<i>Cp38</i>	TGACCTCATCGATCACCTACC	ACGTAGACCTTGCGGAGAAC
<i>E74</i>	AAGAACCCAGCAAACCTGGTG	TTCGGGTTTATCCTGTGAGC
<i>E75</i>	TGCACAAGGTAATGCTGGAG	GTGATCCTGCTTCCTGCTTC
<i>EcR</i>	ACACACGCCTACACATCCAG	GATTTTCGTTTGTCTGAAGTGC
<i>femcoat</i>	AGGCCAAAAAGGAGAAGGAG	GTGACTTCCTCGGACTGCTC
<i>Pxd</i>	ATCAACTGGAGGAGCTGAGG	TTTGGTGAGGTCGACTTGTTG
<i>Pxt</i>	AGTCGAGGGTGGTGTAGTGG	TCGGAGTTGTACACAGGAG
<i>Vm32E</i>	GCTCCAGCTCCAGCTTATTC	TAGTGGGGCACCTGTTCTG
<i>rpl32</i>	CCGCTTCAAGGGACACTATC	GACAATCTCCTTGCGCTTCT
<i>gpdh</i>	GCAAGCTGTCCACCTTCTTC	CGTTGAGCATTTCTTCTCC

**Table A3.2: Genes used for oogenesis study.** Information about each gene tested using qRT-PCR. Oocyte stages in which each gene is reported to be expressed are indicated in “Stages expressed”. A brief description of the relevant gene function is shown in “Gene functions”. **a:** Fold-change reported in [185] (mated/virgin). The first number is the fold change at 24 hours after mating. The number in parentheses is the fold change at 1-3 hours after mating. **b:** Fold-change reported for Acp62F mutant/wild-type in [188].

Gene	Stages expressed	Gene Function	Mated/Virgin fold-change at 24h (1-3h) <sup>a</sup>	Acp62F mutant/wild-type fold change <sup>b</sup>
<i>BR-C / br</i>	5-6	Chorion gene amplification [350, 352]	--	--
<i>Cp15</i>	14	Chorion protein [351]	1.64 (-1.01)	2.42
<i>Cp16</i>	13	Chorion protein [351]	3.3 (-1.55)	1.85
<i>Cp18</i>	14	Chorion protein [351]	3.1 (NA)	--
<i>Cp19</i>	13	Chorion protein [351]	NA (-1.85)	1.53
<i>Cp36</i>	12	Chorion protein [351]	NA (-2.7)	--
<i>Cp38</i>	12	Chorion protein [351]	3.96 (-2.93)	--
<i>E74</i>	early and 8-10	Ecdysone-responsive gene [350]	--	--
<i>E75</i>	early and 8-10	Ecdysone-responsive gene [350, 353]	--	--
<i>EcR</i>	10	Ecdysone receptor [354]	--	--
<i>Femcoat</i>	12	Chorion protein [355]	1.97 (-1.15)	--
<i>Pxd</i>	11	Eggshell crosslinking [356, 357]	--	--
<i>Pxt</i>		Eggshell crosslinking [356, 357]	--	--
<i>Vm32E</i>	10	Vitelline membrane protein [358]	2.12 (-1.16)	--
<i>rpl32</i>	all	housekeeping gene	--	--
<i>gpdh</i>	all	housekeeping gene	--	--



**Figure A3.1: Standard curves from qRT-PCR assays of oogenesis genes.** Standard curves for each primer pair used in this study. Three to six replicate samples were measured at each of four dilution points. Genes with bad standard curves (*br*, *Cp36*, *pxd*, and *Vm32E*) were not used in further analyses.

## RESULTS

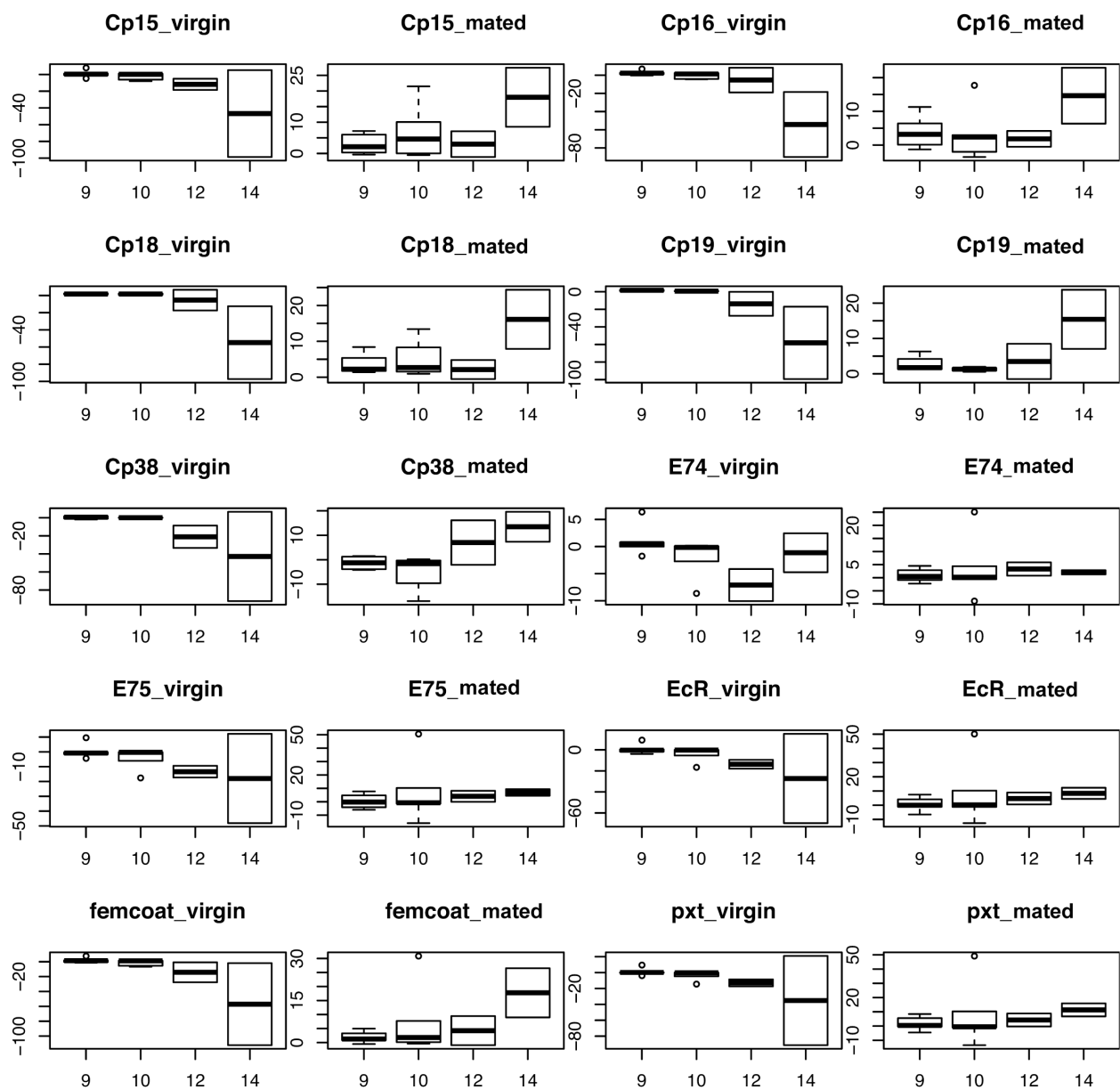
I found that gene expression levels largely agreed with the stage of highest gene expression for a given gene in mated females, though not necessarily in virgin females (Figure A3.2). The stage(s) at which each gene should be most highly expressed is shown in Table A3.1. Mating status did not significantly affect the expression levels of any genes at 10 hours ASM, though I did notice a trend for higher expression of chorion protein genes in stage 13-14 oocytes (Figure A3.3). Similarly, a non-significant trend was observed for Ecdysone-pathway components to be expressed higher in stages 1-12 in mated females versus virgin females (Figure A3.4). I saw no differences, significant or otherwise between stage 1-9 versus stage 10 oocytes. I observed no significant differences or trends between females mated Acp62F null males or controls at either timepoint (data not shown).

**Table A3.3: Results from standard curves.** Technical rep. = number of replicates at each dilution point. Primer pairs with efficiency greater than 1.4 were not used for analysis.

Gene	Slope	Efficiency	Intercept	Technical Reps.
<i>BR-C</i>	-0.204	8.159E4	37.954	3
<i>Cp15</i>	-3.283	1.016	21.925	3
<i>Cp16</i>	-2.934	1.192	29.027	3
<i>Cp18</i>	-3.207	1.05	24.373	3
<i>Cp19</i>	-3.490	0.934	27.26	3
<i>Cp36</i>	-1.754	2.717	24.959	6
<i>Cp38</i>	-3.362	0.984	26.618	3
<i>E74</i>	-3.533	0.919	29.829	4
<i>E75</i>	-3.455	0.947	28.198	4
<i>EcR</i>	-3.790	0.836	30.659	4
<i>femcoat</i>	-3.365	0.982	29.178	3
<i>gpdh</i>	-3.302	1.008	23.185	4
<i>Pxd</i>	-2.290	1.734	37.599	3
<i>Pxt</i>	-3.283	1.016	25.621	4
<i>rpl32</i>	-2.642	1.391	22.88	3
<i>Vm32E</i>	-0.136	2.187E7	38.497	6

**Figure A3.2: qRT-PCR results separated by oocyte stage and mating status.**

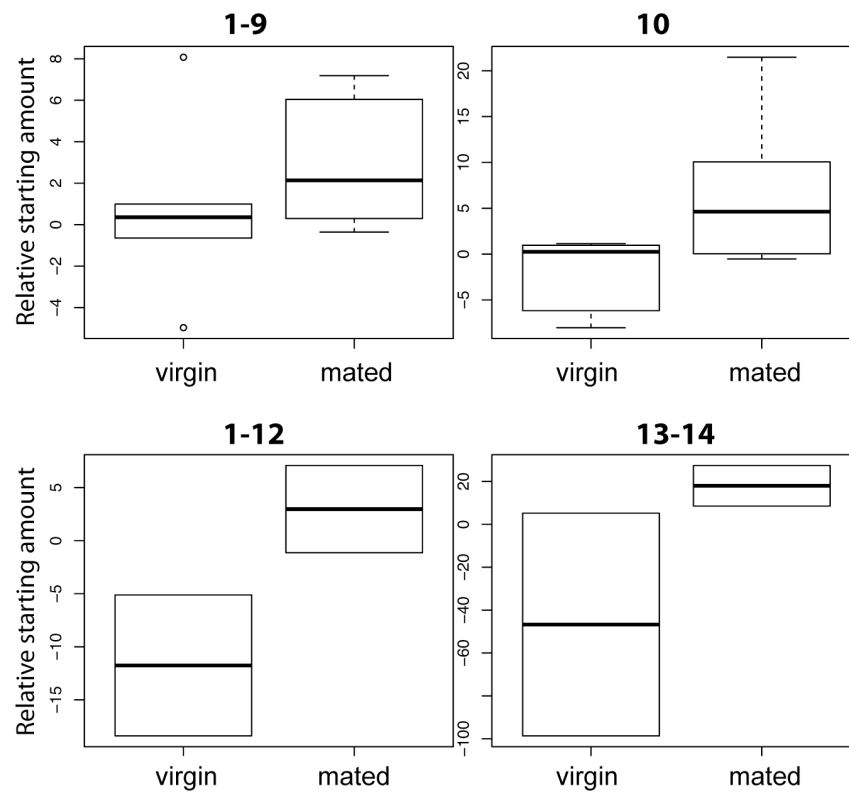
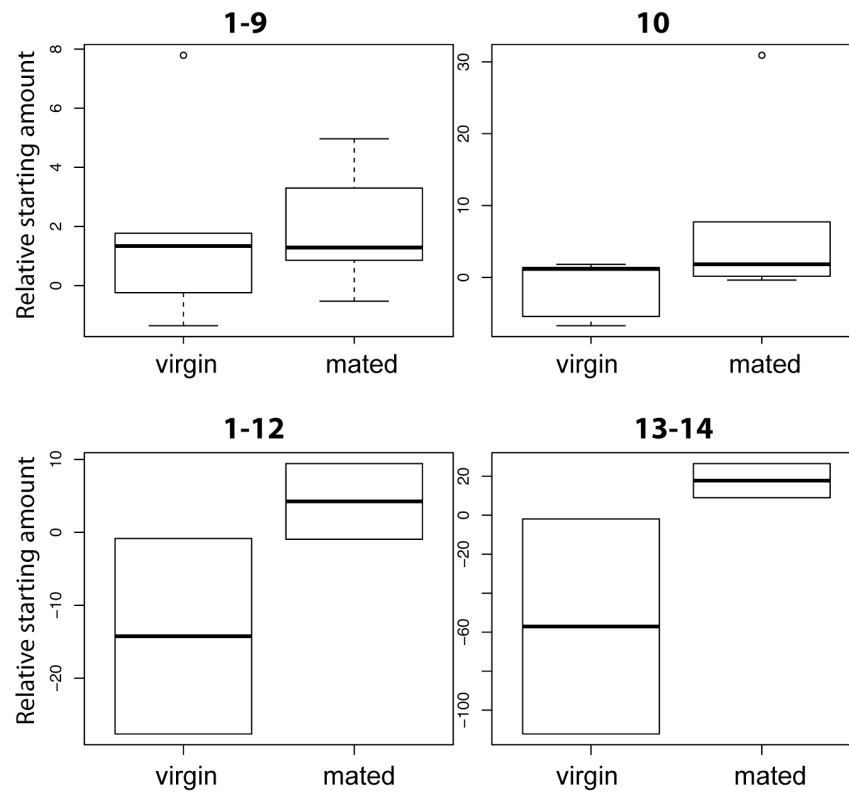
Oocyte stages are given on the X-axis (9 = stages 1-9; 10 = stage 10; 12 = stages 1-12; 14 = stages 13 and 14). Normalized expression values are given on the Y-axis. Samples were prepared from mated females at 10 hours after mating.



**Figure A3.3: Representative plots of chorion protein gene expression, separated by oocyte stage and mating status.**

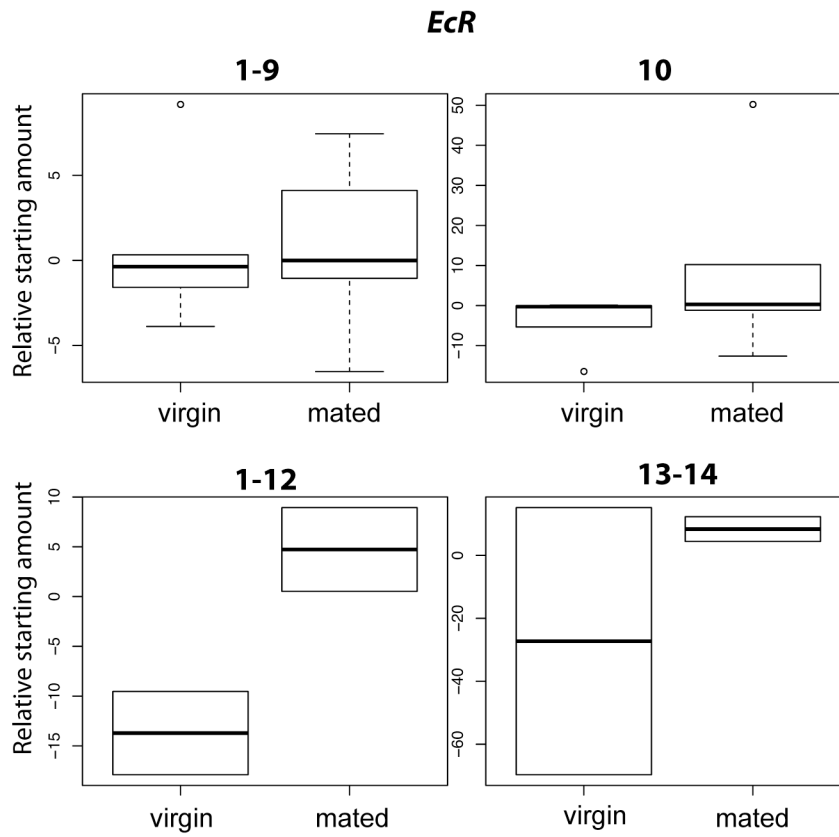
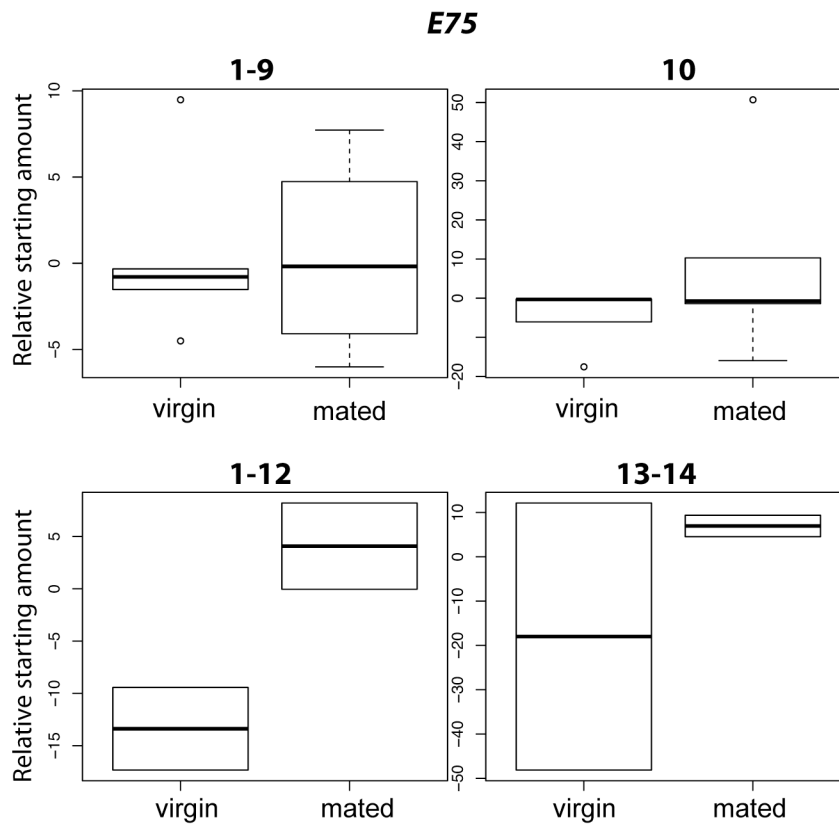
qRT-PCR results from *Cp15* and *Femcoat* are shown as representatives of chorion protein gene expression measured in this study. Oocyte stage pools are shown above each plot. Mating status is given on the X-axis. Normalized gene expression values are given on the Y-axis. Samples were prepared from mated females at 10 hours after mating.



**A*****Cp15*****B*****Femcoat***

**Figure A3.4: Representative plots of ecdysone-responsive gene expression, separated by oocytes stage and mating status.**

qRT-PCR results from *EcR* and *E75* are shown as representatives of ecdysone responsive gene expression measured in this study. Oocyte stage pools are shown above each plot. Mating status is given on the X-axis. Normalized gene expression values are given on the Y-axis.

**A****B**

## CONCLUSIONS

These data do not support the hypothesis that mating induces expression of genes required to lift the stage 9 block to oogenesis that exists in virgin females. There are several possible reasons for this. First, the genes tested here may not be mating-responsive genes. However, this seems unlikely for many of the genes (especially the chorion protein genes), since they were significantly up-regulated in response to mating in previous work [185]. Second, there may have been RNA degradation or sample loss during oocytes dissection, especially in samples with earlier stage oocytes, since these are most difficult to dissect from the ovaries. Most of the differences seen previously [185] were less than 2-fold and a degradation of RNA in some samples could mask any real difference in RNA abundance. Finally, the sample sizes here are small, and there was quite a lot of variation between samples. This is the most likely explanation for the lack of significant results; additional biological replicates might potentially resolve significant differences.

## REFERENCES

1. Poiani A (2006) Complexity of seminal fluid: a review. *Behavioral Ecology and Sociobiology* 60: 289-310.
2. Avila FW, Sirot LK, LaFlamme BA, Rubinstein CD, Wolfner MF (2011) Insect seminal fluid proteins: identification and function. *Annu Rev Entomol* 56: 21-40.
3. Rodríguez-Martínez H, Kvist U, Ernerudh J, Sanz L, Calvete JJ (2011) Seminal plasma proteins: what role do they play? *Am J Reprod Immunol* 66 Suppl 1: 11-22.
4. Kanwar KC, Yanagimachi R, Lopata A (1979) Effects of human seminal plasma on fertilizing capacity of human spermatozoa. *Fertil Steril* 31: 321-327.
5. Osanai M, Chen P (1993) A comparative study on the arginine degradation cascade for sperm maturation of *Bombyx mori* and *Drosophila melanogaster*. *Amino Acids* 5: 341-350.
6. Neubaum DM, Wolfner MF (1999) Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. *Genetics* 153: 845-857.
7. Bloch Qazi MC, Wolfner MF (2003) An early role for the *Drosophila melanogaster* male seminal protein Acp36DE in female sperm storage. *J Exp Biol* 206: 3521-3528.
8. Wong A, Albright S, Giebel J, Ravi Ram K, Ji S, Fiumera A, Wolfner M (2008) A role for Acp29AB, a predicted seminal fluid lectin, in female sperm storage in *Drosophila melanogaster*. *Genetics*.
9. King M, Eubel H, Millar AH, Baer B (2011) Proteins within the seminal fluid are crucial to keep sperm viable in the honeybee *Apis mellifera*. *J Insect Physiol* 57: 409-414.
10. Lu C-H, Lee RK-K, Hwu Y-M, Chu S-L, Chen Y-J, Chang W-C, Lin S-P, Li S-H (2011) SERPINE2, a serine protease inhibitor extensively expressed in adult male mouse reproductive tissues, may serve as a murine sperm decapacitation factor. *Biol Reprod* 84: 514-525.
11. Smith JR, Stanfield GM (2011) TRY-5 is a sperm-activating protease in *Caenorhabditis elegans* seminal fluid. *PLoS Genet* 7: e1002375.

12. Ou C-M, Tang J-B, Huang M-S, Sudhakar Gandhi PS, Geetha S, Li S-H, Chen Y-H (2012) The mode of reproductive-derived Spink (serine protease inhibitor Kazal-type) action in the modulation of mammalian sperm activity. *Int J Androl* 35: 52-62.
13. Zhao Y, Sun W, Zhang P, Chi H, Zhang M-J, Song C-Q, Ma X, Shang Y, Wang B, Hu Y, Hao Z, Hühmer AF, Meng F, L'hernault SW, He S-M, Dong M-Q, Miao L (2012) Nematode sperm maturation triggered by protease involves sperm-secreted serine protease inhibitor (Serp). *Proc Natl Acad Sci USA* 109: 1542-1547.
14. Daimon E, Wada Y (2005) Role of neutrophils in matrix metalloproteinase activity in the preimplantation mouse uterus. *Biol Reprod* 73: 163-171.
15. Moldenhauer LM, Diener KR, Thring DM, Brown MP, Hayball JD, Robertson SA (2009) Cross-presentation of male seminal fluid antigens elicits T cell activation to initiate the female immune response to pregnancy. *J Immunol* 182: 8080-8093.
16. Guerin LR, Moldenhauer LM, Prins JR, Bromfield JJ, Hayball JD, Robertson SA (2011) Seminal Fluid Regulates Accumulation of FOXP3+ Regulatory T Cells in the Preimplantation Mouse Uterus Through Expanding the FOXP3+ Cell Pool and CCL19-Mediated Recruitment. *Biol Reprod*.
17. Ziecik AJ, Wacławik A, Kaczmarek MM, Blitek A, Jalali BM, Andronowska A (2011) Mechanisms for the establishment of pregnancy in the pig. *Reprod Domest Anim* 46 Suppl 3: 31-41.
18. Doyle U, Sampson N, Zenzmaier C, Schwärzler P, Berger P (2012) Seminal plasma enhances and accelerates progesterone-induced decidualisation of human endometrial stromal cells. *Reprod Fertil Dev* 24: 517-522.
19. Domanitskaya EV, Liu H, Chen S, Kubli E (2007) The hydroxyproline motif of male sex peptide elicits the innate immune response in *Drosophila* females. *FEBS J* 274: 5659-5668.
20. Heifetz Y, Lung O, Frongillo EA, Wolfner MF (2000) The *Drosophila* seminal fluid protein Acp26Aa stimulates release of oocytes by the ovary. *Curr Biol* 10: 99-102.
21. Ratto MH, Huanca W, Adams GP (2010) Ovulation-inducing factor: a protein component of llama seminal plasma. *Reprod Biol Endocrinol* 8: 44.

22. Marshall JL, Huestis DL, Hiromasa Y, Wheeler S, Oppert C, Marshall SA, Tomich JM, Oppert B (2009) Identification, RNAi knockdown, and functional analysis of an ejaculate protein that mediates a postmating, prezygotic phenotype in a cricket. *PLoS ONE* 4: e7537.
23. Xu J, Wang Q (2011) Seminal fluid reduces female longevity and stimulates egg production and sperm trigger oviposition in a moth. *J Insect Physiol* 57: 385-390.
24. Gillott C (2003) Male accessory gland secretions: modulators of female reproductive physiology and behavior. *Annu Rev Entomol* 48: 163-184.
25. Sirot LK, LaFlamme BA, Sitnik JL, Rubinstein CD, Avila FW, Chow CY, Wolfner MF (2009) Molecular social interactions: *Drosophila melanogaster* seminal fluid proteins as a case study. *Adv Genet* 68: 23-56.
26. Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro C (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc Natl Acad Sci USA* 98: 7375-7379.
27. Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. *Nat Rev Genet* 3: 137-144.
28. Clark NL, Swanson WJ (2005) Pervasive adaptive evolution in primate seminal proteins. *PLoS Genet* 1: e35.
29. Clark NL, Aagaard JE, Swanson WJ (2006) Evolution of reproductive proteins from animals and plants. *Reproduction* 131: 11-22.
30. Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Kristipati RR, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, Singh RS (2007) Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* 177: 1321-1335.
31. Mueller JL, Ripoll DR, Aquadro C, Wolfner MF (2004) Comparative structural modeling and inference of conserved protein classes in *Drosophila* seminal fluid. *Proc Natl Acad Sci USA* 101: 13542-13547.
32. Udvardy A (1996) The role of controlled proteolysis in cell-cycle regulation. *Eur J Biochem* 240: 307-313.

33. Heutinck KM, ten Berge IJM, Hack CE, Hamann J, Rowshani AT (2010) Serine proteases of the human immune system in health and disease. *Mol Immunol* 47: 1943-1955.
34. Di Cera E, editor (2011) *Proteases in Health and Disease*. 1 ed. San Diego, CA: Elsevier, Inc. 328 p.
35. Behrendt N, editor (2012) *Matrix Proteases in Health and Disease*. Weinheim, Germany: Wiley-VCH. 416 p.
36. Silverman GA, Whisstock JC, Bottomley SP, Huntington JA, Kaiserman D, Luke CJ, Pak SC, Reichhart J-M, Bird PI (2010) Serpins flex their muscle: I. Putting the clamps on proteolysis in diverse biological systems. *J Biol Chem* 285: 24299-24305.
37. Farady CJ, Craik CS (2010) Mechanisms of macromolecular protease inhibitors. *Chembiochem* 11: 2341-2346.
38. Polgár L (1989) *Mechanisms of Protease Action*: Taylor & Francis, Inc. 223 p.
39. Ross J, Jiang H, Kanost MR, Wang Y (2003) Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. *Gene* 304: 117-131.
40. Shah PK, Tripathi LP, Jensen LJ, Gahnim M, Mason C, Furlong EE, Rodrigues V, White KP, Bork P, Sowdhamini R (2008) Enhanced function annotations for *Drosophila* serine proteases: a case study for systematic annotation of multi-member gene families. *Gene* 407: 199-215.
41. Page MJ, Di Cera E (2008) Serine peptidases: classification, structure and function. *Cell Mol Life Sci* 65: 1220-1236.
42. Zitka O, Kukacka J, Krizkova S, Huska D, Adam V, Masarik M, Prusa R, Kizek R (2010) Matrix metalloproteinases. *Curr Med Chem* 17: 3751-3768.
43. Bond JS, Beynon RJ (1995) The astacin family of metalloendopeptidases. *Protein Sci* 4: 1247-1261.
44. Domsalla A, Melzig MF (2008) Occurrence and properties of proteases in plant latices. *Planta Med* 74: 699-711.



45. Chapman HA, Riese RJ, Shi GP (1997) Emerging roles for cysteine proteases in human biology. *Annu Rev Physiol* 59: 63-88.
46. Turk V, Stoka V, Vasiljeva O, Renko M, Sun T, Turk B, Turk D (2012) Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim Biophys Acta* 1824: 68-88.
47. Utleg AG, Yi EC, Xie T, Shannon P, White JT, Goodlett DR, Hood L, Lin B (2003) Proteomic analysis of human prostasomes. *Prostate* 56: 150-161.
48. Fung KYC, Glode LM, Green S, Duncan MW (2004) A comprehensive characterization of the peptide and protein constituents of human seminal fluid. *Prostate* 61: 171-181.
49. Szecsi PB (1992) The aspartic proteases. *Scand J Clin Lab Invest Suppl* 210: 5-22.
50. Gubb D, Sanz-Parra A, Barcena L, Troxler L, Fullaondo A (2010) Protease inhibitors and proteolytic signalling cascades in insects. *Biochimie* 92: 1749-1759.
51. Duncan MW, Thompson HS (2007) Proteomics of semen and its constituents. *Proteomics Clin Appl* 1: 861-875.
52. Pilch B, Mann M (2006) Large-scale and high-confidence proteomic analysis of human seminal plasma. *Genome Biol* 7: R40.
53. Yousef GM, Diamandis EP (2001) The new human tissue kallikrein gene family: structure, function, and association to disease. *Endocr Rev* 22: 184-204.
54. Clauss A, Persson M, Lilja H, Lundwall Å (2011) Three genes expressing Kunitz domains in the epididymis are related to genes of WFDC-type protease inhibitors and semen coagulum proteins in spite of lacking similarity between their protein products. *BMC Biochem* 12: 55.
55. Mueller JL, Kristipati RR, McGraw LA, Bloch Qazi MC, Siggia ED, Clark AG, Aquadro C, Wolfner MF (2005) Cross-species comparison of *Drosophila* male accessory gland protein genes. *Genetics* 171: 131-143.
56. Walker MJ, Rylett CM, Keen JN, Audsley N, Sajid M, Shirras AD, Isaac RE (2006) Proteomic identification of *Drosophila melanogaster* male accessory gland proteins,

including a pro-cathepsin and a soluble gamma-glutamyl transpeptidase. *Proteome science* 4: 9.

57. Findlay GD, Yi X, Maccoss MJ, Swanson WJ (2008) Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol* 6: e178.
58. Findlay GD, Maccoss MJ, Swanson WJ (2009) Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in *Drosophila*. *Genome Res* 19: 886-896.
59. Takemori N, Yamamoto M-T (2009) Proteome mapping of the *Drosophila melanogaster* male reproductive system. *Proteomics* 9: 2484-2493.
60. Ayroles JF, LaFlamme BA, Stone EA, Wolfner MF, Mackay TFC (2011) Functional genome annotation of *Drosophila* seminal fluid proteins using transcriptional genetic networks. *Genet Res (Camb)* 93: 387-395.
61. Dorus S, Busby SA, Gerike U, Shabanowitz J, Hunt DF, Karr TL (2006) Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. *Nat Genet* 38: 1440-1445.
62. Chintapalli VR, Wang J, Dow JAT (2007) Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet* 39: 715-720.
63. LaFlamme BA, Ravi Ram K, Wolfner MF (2012) The *Drosophila melanogaster* Seminal Fluid Protease “Seminase” Regulates Proteolytic and Post-Mating Reproductive Processes. *PLoS Genet* 8: e1002435.
64. Ellis LL, Carney GE (2010) Mating alters gene expression patterns in *Drosophila melanogaster* male heads. *BMC Genomics* 11: 558.
65. Mueller JL, Page JL, Wolfner MF (2007) An ectopic expression screen reveals the protective and toxic effects of *Drosophila* seminal fluid proteins. *Genetics* 175: 777-783.
66. Wong A, Turchin MC, Wolfner MF, Aquadro C (2008) Evidence for positive selection on *Drosophila melanogaster* seminal fluid protease homologs. *Mol Biol Evol* 25: 497-506.
67. Ravi Ram K, Wolfner MF (2007) Sustained post-mating response in *Drosophila*

- melanogaster requires multiple seminal fluid proteins. PLoS Genet 3: e238.
68. Ravi Ram K, Wolfner MF (2009) A network of interactions among seminal proteins underlies the long-term postmating response in *Drosophila*. Proc Natl Acad Sci USA 106: 15384-15389.
  69. Avila FW, Ravi Ram K, Bloch Qazi MC, Wolfner MF (2010) Sex peptide is required for the efficient release of stored sperm in mated *Drosophila* females. Genetics 186: 595-600.
  70. Kelleher E, Pennington J (2009) Protease Gene Duplication and Proteolytic Activity in *Drosophila* Female Reproductive Tracts. Mol Biol Evol 26: 2125-2134.
  71. Garrett M, Fullaondo A, Troxler L, Micklem G, Gubb D (2009) Identification and analysis of serpin-family genes by homology and synteny across the 12 sequenced *Drosophilid* genomes. BMC Genomics 10: 489.
  72. Richer MJ, Keays CA, Waterhouse J, Minhas J, Hashimoto C, Jean F (2004) The Spn4 gene of *Drosophila* encodes a potent furin-directed secretory pathway serpin. Proc Natl Acad Sci USA 101: 10560-10565.
  73. Reichhart JM, Gubb D, Leclerc V (2011) The *Drosophila* serpins: multiple functions in immunity and morphogenesis. Meth Enzymol 499: 205-225.
  74. Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RRH, Mackay TFC (2009) Systems genetics of complex traits in *Drosophila melanogaster*. Nat Genet 41: 299-307.
  75. Begun DJ, Lindfors HA (2005) Rapid evolution of genomic Acp complement in the melanogaster subgroup of *Drosophila*. Mol Biol Evol 22: 2010-2021.
  76. Lung O, Wolfner MF (1999) *Drosophila* seminal fluid proteins enter the circulatory system of the mated female fly by crossing the posterior vaginal wall. Insect Biochem Mol Biol 29: 1043-1052.
  77. Lung O, Tram U, Finnerty CM, Eipper-Mains MA, Kalb JM, Wolfner MF (2002) The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. Genetics 160: 211-224.

78. Ravi Ram K, Ji S, Wolfner MF (2005) Fates and targets of male accessory gland proteins in mated female *Drosophila melanogaster*. *Insect Biochem Mol Biol* 35: 1059-1071.
79. Fiumera AC, Dumont BL, Clark AG (2007) Associations between sperm competition and natural variation in male reproductive genes on the third chromosome of *Drosophila melanogaster*. *Genetics* 176: 1245-1260.
80. Mueller JL, Linklater JR, Kristipati RR, Chapman T, Wolfner MF (2008) Targeted gene deletion and phenotypic analysis of the *Drosophila melanogaster* seminal fluid protease inhibitor Acp62F. *Genetics* 178: 1605-1614.
81. Ravi Ram K, Sirot LK, Wolfner MF (2006) Predicted seminal astacin-like protease is required for processing of reproductive proteins in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 103: 18674-18679.
82. Greenspan L, Clark AG (2011) Associations between Variation in X Chromosome Male Reproductive Genes and Sperm Competitive Ability in *Drosophila melanogaster*. *Int J Evol Biol* 2011: 214280.
83. Kelleher ES, Watts TD, LaFlamme BA, Haynes PA, Markow TA (2009) Proteomic analysis of *Drosophila mojavensis* male accessory glands suggests novel classes of seminal fluid proteins. *Insect Biochem Mol Biol* 39: 366-371.
84. Wagstaff BJ, Begun DJ (2005) Molecular population genetics of accessory gland protein genes and testis-expressed genes in *Drosophila mojavensis* and *D. arizonae*. *Genetics* 171: 1083-1101.
85. Dottorini T, Nicolaides L, Ranson H, Rogers DW, Crisanti A, Catteruccia F (2007) A genome-wide analysis in *Anopheles gambiae* mosquitoes reveals 46 male accessory gland genes, possible modulators of female behavior. *Proc Natl Acad Sci USA* 104: 16215-16220.
86. Azevedo RVDM, Dias DBS, Bretãs JAC, Mazzoni CJ, Souza NA, Albano RM, Wagner G, Davila AMR, Peixoto AA (2012) The Transcriptome of *Lutzomyia longipalpis* (Diptera: Psychodidae) Male Reproductive Organs. *PLoS ONE* 7: e34495.
87. Davies SJ, Chapman T (2006) Identification of genes expressed in the accessory glands of male Mediterranean Fruit Flies (*Ceratitis capitata*). *Insect Biochem Mol Biol* 36: 846-856.

88. Baer B, Heazlewood JL, Taylor NL, Eubel H, Millar AH (2009) The seminal fluid proteome of the honeybee *Apis mellifera*. *Proteomics* 9: 2085-2097.
89. Walters JR, Harrison RG (2008) EST analysis of male accessory glands from *Heliconius* butterflies with divergent mating systems. *BMC Genomics* 9: 592.
90. Walters JR, Harrison RG (2010) Combined EST and proteomic analysis identifies rapidly evolving seminal fluid proteins in *Heliconius* butterflies. *Mol Biol Evol* 27: 2000-2013.
91. Nagaoka S, Kato K, Takata Y, Kamei K (2012) Identification of the sperm-activating factor initiatorin, a prostatic endopeptidase of the silkworm, *Bombyx mori*. *Insect Biochem Mol Biol*.
92. South A, Sirot LK, Lewis SM (2011) Identification of predicted seminal fluid proteins in *Tribolium castaneum*. *Insect Mol Biol* 20: 447-456.
93. Andrés JA, Maroja LS, Bogdanowicz SM, Swanson WJ, Harrison RG (2006) Molecular evolution of seminal proteins in field crickets. *Mol Biol Evol* 23: 1574-1584.
94. Weiss BL, Stepczynski JM, Wong P, Kaufman WR (2002) Identification and characterization of genes differentially expressed in the testis/vas deferens of the fed male tick, *Amblyomma hebraeum*. *Insect Biochem Mol Biol* 32: 785-793.
95. Dean MD, Findlay GD, Hoopmann MR, Wu CC, Maccoss MJ, Swanson WJ, Nachman MW (2011) Identification of ejaculated proteins in the house mouse (*Mus domesticus*) via isotopic labeling. *BMC Genomics* 12: 306.
96. Claydon AJ, Ramm SA, Pennington A, Hurst JL, Stockley P, Beynon RJ (2012) Heterogenous turnover of sperm and seminal vesicle proteins in the mouse revealed by dynamic metabolic labelling. *Mol Cell Proteomics*.
97. Kelly VC, Kuy S, Palmer DJ, Xu Z, Davis SR, Cooper GJ (2006) Characterization of bovine seminal plasma by proteomics. *Proteomics* 6: 5826-5833.
98. Moura AA, Souza CE, Stanley BA, Chapman DA, Killian GJ (2010) Proteomics of cauda epididymal fluid from mature Holstein bulls. *J Proteomics* 73: 2006-2020.
99. Novak S, Smith TA, Paradis F, Burwash L, Dyck MK, Foxcroft GR, Dixon WT (2010)

- Biomarkers of in vivo fertility in sperm and seminal plasma of fertile stallions. *Theriogenology* 74: 956-967.
100. Wojtczak M, Dietrich GJ, Ciereszko A (2005) Transferrin and antiproteases are major proteins of common carp seminal plasma. *Fish Shellfish Immunol* 19: 387-391.
  101. Mommens M, Wojtczak M, Ciereszko A, Babiak I (2008) Seminal plasma proteins of Atlantic halibut (*Hippoglossus hippoglossus* L.). *Fish Physiol Biochem* 34: 349-355.
  102. Thurston RJ, Korn N, Froman DP, Bodine AB (1993) Proteolytic enzymes in seminal plasma of domestic turkey (*Meleagris gallopavo*). *Biol Reprod* 48: 393-402.
  103. Kotłowska M, Kowalski R, Glogowski J, Jankowski J, Ciereszko A (2005) Gelatinases and serine proteinase inhibitors of seminal plasma and the reproductive tract of turkey (*Meleagris gallopavo*). *Theriogenology* 63: 1667-1681.
  104. Marshall JL, Huestis DL, Garcia C, Hiromasa Y, Wheeler S, Noh S, Tomich JM, Howard DJ (2011) Comparative proteomics uncovers the signature of natural selection acting on the ejaculate proteomes of two cricket species isolated by postmating, prezygotic phenotypes. *Mol Biol Evol* 28: 423-435.
  105. Herndon LA, Wolfner MF (1995) A *Drosophila* seminal fluid protein, Acp26Aa, stimulates egg laying in females for 1 day after mating. *Proc Natl Acad Sci USA* 92: 10114-10118.
  106. Park M, Wolfner MF (1995) Male and female cooperate in the prohormone-like processing of a *Drosophila melanogaster* seminal fluid protein. *Dev Biol* 171: 694-702.
  107. Sirot LK, Findlay GD, Sitnik JL, Frasher D, Avila FW, Wolfner M (2012) A family of female- and male-derived reproductive tract proteins in *Drosophila* that appears to have evolved through gene duplication. In prep.
  108. Heifetz Y, Vandenberg LN, Cohn HI, Wolfner MF (2005) Two cleavage products of the *Drosophila* accessory gland protein ovulin can independently induce ovulation. *Proc Natl Acad Sci USA* 102: 743-748.
  109. Lilja H, Oldbring J, Rannevik G, Laurell CB (1987) Seminal vesicle-secreted proteins and their reactions during gelation and liquefaction of human semen. *J Clin Invest* 80: 281-285.

110. Lilja H, Abrahamsson PA, Lundwall A (1989) Semenogelin, the predominant protein in human semen. Primary structure and identification of closely related proteins in the male accessory sex glands and on the spermatozoa. *J Biol Chem* 264: 1894-1900.
111. Borogño CA, Diamandis EP (2004) The emerging roles of human tissue kallikreins in cancer. *Nat Rev Cancer* 4: 876-890.
112. Pampalakis G, Sotiropoulou G (2007) Tissue kallikrein proteolytic cascade pathways in normal physiology and cancer. *Biochim Biophys Acta* 1776: 22-31.
113. Matsuda Y, Oshio S, Yazaki T, Umeda T, Akihama S (1994) The effect of some proteinase inhibitors on liquefaction of human semen. *Hum Reprod* 9: 664-668.
114. Suzuki K, Kise H, Nishioka J, Hayashi T (2007) The interaction among protein C inhibitor, prostate-specific antigen, and the semenogelin system. *Semin Thromb Hemost* 33: 46-52.
115. Williams SA, Xu Y, De Marzo AM, Isaacs JT, Denmeade SR (2010) Prostate-specific antigen (PSA) is activated by KLK2 in prostate cancer ex vivo models and in prostate-targeted PSA/KLK2 double transgenic mice. *Prostate* 70: 788-796.
116. Lundwall A, Brattsand M (2008) Kallikrein-related peptidases. *Cell Mol Life Sci* 65: 2019-2038.
117. Emami N, Diamandis EP (2008) Human kallikrein-related peptidase 14 (KLK14) is a new activator component of the KLK proteolytic cascade. Possible function in seminal plasma and skin. *J Biol Chem* 283: 3031-3041.
118. Luo L-Y, Shan SJC, Elliott MB, Soosaipillai A, Diamandis EP (2006) Purification and characterization of human kallikrein 11, a candidate prostate and ovarian cancer biomarker, from seminal plasma. *Clin Cancer Res* 12: 742-750.
119. Lwaleed BA, Greenfield R, Stewart A, Birch B, Cooper AJ (2004) Seminal clotting and fibrinolytic balance: a possible physiological role in the male reproductive system. *Thromb Haemost* 92: 752-766.
120. Lwaleed BA, Greenfield RS, Birch BR, Cooper AJ (2005) Does human semen contain a functional haemostatic system? A possible role for Tissue Factor Pathway Inhibitor in fertility through semen liquefaction. *Thromb Haemost* 93: 847-852.

121. Lwaleed BA, Goyal A, Delves G, Gossai S, Greenfield RS, Cooper AJ (2007) Seminal factor VII and factor VIIa: supporting evidence for the presence of an active tissue factor-dependent coagulation pathway in human semen. *Int J Androl* 30: 543-549.
122. Fernández JA, Heeb MJ (2007) Role of protein C inhibitor and tissue factor in fertilization. *Semin Thromb Hemost* 33: 13-20.
123. Fernández JA, Heeb MJ, Radtke KP, Griffin JH (1997) Potent blood coagulant activity of human semen due to prostatic-bound tissue factor. *Biol Reprod* 56: 757-763.
124. Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT (2004) Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nat Genet* 36: 1326-1329.
125. Carnahan SJ, Jensen-Seaman MI (2008) Hominoid seminal protein evolution and ancestral mating behavior. *Am J Primatol* 70: 939-948.
126. Jensen-Seaman MI, Li W-H (2003) Evolution of the hominoid semenogelin genes, the major proteins of ejaculated semen. *J Mol Evol* 57: 261-270.
127. Kingan SB, Tatar M, Rand DM (2003) Reduced polymorphism in the chimpanzee semen coagulating protein, semenogelin I. *J Mol Evol* 57: 159-169.
128. Tseng H-C, Tang J-B, Gandhi PSS, Luo C-W, Ou C-M, Tseng C-J, Lin H-J, Chen Y-H (2012) Mutual adaptation between mouse transglutaminase 4 and its native substrates in the formation of copulatory plug. *Amino Acids* 42: 951-960.
129. Bretman A, Lawniczak MKN, Boone J, Chapman T (2010) A mating plug protein reduces early female remating in *Drosophila melanogaster*. *J Insect Physiol* 56: 107-113.
130. Murer V, Spetz JF, Hengst U, Altrogge LM, de Agostini A, Monard D (2001) Male fertility defects in mice lacking the serine protease inhibitor protease nexin-1. *Proc Natl Acad Sci USA* 98: 3029-3033.
131. Rogers DW, Baldini F, Battaglia F, Panico M, Dell A, Morris HR, Catteruccia F (2009) Transglutaminase-mediated semen coagulation controls sperm storage in the malaria mosquito. *PLoS Biol* 7: e1000272.



132. Holt WV (2011) Mechanisms of sperm storage in the female reproductive tract: an interspecies comparison. *Reprod Domest Anim* 46 Suppl 2: 68-74.
133. Suarez SS (2008) Regulation of sperm storage and movement in the mammalian oviduct. *Int J Dev Biol* 52: 455-462.
134. Hung PH, Suarez SS (2010) Regulation of sperm storage and movement in the ruminant oviduct. *Society of Reproduction and Fertility supplement* 67: 257-266.
135. Avila FW, Wolfner MF (2009) Acp36DE is required for uterine conformational changes in mated *Drosophila* females. *Proc Natl Acad Sci USA* 106: 15796-15800.
136. Chapman T, Neubaum DM, Wolfner MF, Partridge L (2000) The role of male accessory gland protein Acp36DE in sperm competition in *Drosophila melanogaster*. *Proc Biol Sci* 267: 1097-1105.
137. Lung O, Wolfner MF (2001) Identification and characterization of the major *Drosophila melanogaster* mating plug protein. *Insect Biochem Mol Biol* 31: 543-551.
138. Parker GA, Pizzari T (2010) Sperm competition and ejaculate economics. *Biological reviews of the Cambridge Philosophical Society* 85: 897-934.
139. Sirot LK, Wolfner MF, Wigby S (2011) Protein-specific manipulation of ejaculate composition in response to female mating status in *Drosophila melanogaster*. *Proc Natl Acad Sci USA*.
140. Dacheux J-L, Gatti JL, Dacheux F (2003) Contribution of epididymal secretory proteins for spermatozoa maturation. *Microsc Res Tech* 61: 7-17.
141. Guyonnet B, Dacheux F, Dacheux J-L, Gatti J-L (2011) The epididymal transcriptome and proteome provide some insights into new epididymal regulations. *J Androl* 32: 651-664.
142. Lee B, Park I, Jin S, Choi H, Kwon JT, Kim J, Jeong J, Cho B-N, Eddy EM, Cho C (2011) Impaired spermatogenesis and fertility in mice carrying a mutation in the *Spink2* gene expressed predominantly in testes. *J Biol Chem* 286: 29108-29117.
143. Zalazar L, Saez Lancellotti TE, Clementi M, Lombardo C, Lamattina L, De Castro R, Fornés MW, Cesari A (2012) *SPINK3* modulates mouse sperm physiology through the

- reduction of nitric oxide level independently of its trypsin inhibitory activity. *Reproduction* 143: 281-295.
144. de Lamirande E, Lamothe G (2009) Reactive oxygen-induced reactive oxygen formation during human sperm capacitation. *Free Radic Biol Med* 46: 502-510.
  145. Roessner C, Paasch U, Glander H-J, Grunewald S (2010) Activity of nitric oxide synthase in mature and immature human spermatozoa. *Andrologia* 42: 132-137.
  146. Lin M, Lee R, Hwu Y, Lu C, Chu S, Chen Y, Chang W, Li S (2008) SPINKL, a Kazal-type serine protease inhibitor-like protein purified from mouse seminal vesicle fluid, is able to inhibit sperm capacitation. *Reproduction*.
  147. Stanfield GM, Villeneuve AM (2006) Regulation of sperm activation by SWM-1 is required for reproductive success of *C. elegans* males. *Curr Biol* 16: 252-263.
  148. Ward S, Carrel JS (1979) Fertilization and sperm competition in the nematode *Caenorhabditis elegans*. *Dev Biol* 73: 304-321.
  149. Friedländer M, Jeshtadi A, Reynolds S (2001) The structural mechanism of trypsin-induced intrinsic motility in *Manduca sexta* spermatozoa in vitro. *J Insect Physiol* 47: 245-255.
  150. Aigaki T, Kasuga H, Nagaoka S, Osanai M (1994) Purification and partial amino acid sequence of initiatorin, a prostatic endopeptidase of the silkworm, *Bombyx mori*. *Insect Biochem Mol Biol* 24: 969-975.
  151. Miyata H, Thaler CD, Haimo LT, Cardullo RA (2012) Protease activation and the signal transduction pathway regulating motility in sperm from the water strider *Aquarius remigis*. *Cytoskeleton* (Hoboken, NJ).
  152. Ding X, Zhang J, Bian Z, Xia Y, Lu C, Gu A, Li Y, Song L, Wang S, Wang X (2010) Variants in the Eppin gene show association with semen quality in Han-Chinese population. *Reprod Biomed Online* 20: 125-131.
  153. O'Rand MG, Widgren EE, Hamil KG, Silva EJ, Richardson RT (2011) Functional studies of eppin. *Biochem Soc Trans* 39: 1447-1449.
  154. Mitra A, Richardson RT, O'Rand MG (2010) Analysis of recombinant human semenogelin

- as an inhibitor of human sperm motility. *Biol Reprod* 82: 489-496.
155. O'Rand MG, Widgren EE, Sivashanmugam P, Richardson RT, Hall SH, French FS, VandeVoort CA, Ramachandra SG, Ramesh V, Jagannadha Rao A (2004) Reversible immunocontraception in male monkeys immunized with eppin. *Science* 306: 1189-1190.
  156. O'Rand MG, Widgren EE, Beyler S, Richardson RT (2009) Inhibition of human sperm motility by contraceptive anti-eppin antibodies from infertile male monkeys: effect on cyclic adenosine monophosphate. *Biol Reprod* 80: 279-285.
  157. Tentes I, Asimakopoulos B, Mourvati E, Diedrich K, Al-Hasani S, Nikolettos N (2007) Matrix metalloproteinase (MMP)-2 and MMP-9 in seminal plasma. *J Assist Reprod Genet* 24: 278-281.
  158. Saengsoi W, Shia W-Y, Shyu C-L, Wu J-T, Warinrak C, Lee W-M, Cheng F-P (2011) Detection of matrix metalloproteinase (MMP)-2 and MMP-9 in canine seminal plasma. *Anim Reprod Sci* 127: 114-119.
  159. Robertson SA (2005) Seminal plasma and male factor signalling in the female reproductive tract. *Cell Tissue Res* 322: 43-52.
  160. Emami N, Diamandis EP (2010) Potential role of multiple members of the kallikrein-related peptidase family of serine proteases in activating latent TGF beta 1 in semen. *Biol Chem* 391: 85-95.
  161. Nocera M, Chu TM (1993) Transforming growth factor beta as an immunosuppressive protein in human seminal plasma. *Am J Reprod Immunol* 30: 1-8.
  162. Nocera M, Chu TM (1995) Characterization of latent transforming growth factor-beta from human seminal plasma. *Am J Reprod Immunol* 33: 282-291.
  163. Robertson SA, Ingman WV, O'Leary S, Sharkey DJ, Tremellen KP (2002) Transforming growth factor beta--a mediator of immune deviation in seminal plasma. *J Reprod Immunol* 57: 109-128.
  164. Mårdh PA, Colleen S (1975) Antimicrobial activity of human seminal fluid. *Scand J Urol Nephrol* 9: 17-23.

165. Marchini D, Marri L, Rosetto M, Manetti AG, Dallai R (1997) Presence of antibacterial peptides on the laid egg chorion of the medfly *Ceratitis capitata*. *Biochem Biophys Res Commun* 240: 657-663.
166. Lung O, Kuo L, Wolfner MF (2001) *Drosophila* males transfer antibacterial proteins from their accessory gland and ejaculatory duct to their mates. *J Insect Physiol* 47: 617-622.
167. Edström AML, Malm J, Frohm B, Martellini JA, Giwercman A, Mörgelin M, Cole AM, Sørensen OE (2008) The major bactericidal activity of human seminal plasma is zinc-dependent and derived from fragmentation of the semenogelins. *J Immunol* 181: 3413-3421.
168. Short SM, Wolfner MF, Lazzaro BP (2012) Female *Drosophila melanogaster* suffer reduced defense against infection due to seminal fluid components. *J Insect Physiol*.
169. Roan NR, Müller JA, Liu H, Chu S, Arnold F, Stürzel CM, Walther P, Dong M, Witkowska HE, Kirchhoff F, Münch J, Greene WC (2011) Peptides released by physiological cleavage of semen coagulum proteins form amyloids that enhance HIV infection. *Cell Host Microbe* 10: 541-550.
170. Weidinger S, Ring J, Köhn FM (2005) IgE-mediated allergy against human seminal plasma. *Chem Immunol Allergy* 88: 128-138.
171. Waldinger MD, Meinardi MMHM, Zwinderman AH, Schweitzer DH (2011) Postorgasmic Illness Syndrome (POIS) in 45 Dutch Caucasian Males: Clinical Characteristics and Evidence for an Immunogenic Pathogenesis (Part 1). *The journal of sexual medicine* 8: 1164-1170.
172. Chapman T, Liddle LF, Kalb JM, Wolfner MF, Partridge L (1995) Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. *Nature* 373: 241-244.
173. Wigby S, Chapman T (2005) Sex peptide causes mating costs in female *Drosophila melanogaster*. *Curr Biol* 15: 316-321.
174. Agard NJ, Wells JA (2009) Methods for the proteomic identification of protease substrates. *Curr Opin Chem Biol* 13: 503-509.
175. Dix MM, Simon GM, Cravatt BF (2008) Global mapping of the topography and magnitude

- of proteolytic events in apoptosis. *Cell* 134: 679-691.
176. Schilling O, Overall CM (2008) Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol* 26: 685-694.
177. Prudova A, Auf dem Keller U, Butler GS, Overall CM (2010) Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. *Mol Cell Proteomics* 9: 894-911.
178. Becker-Pauly C, Barré O, Schilling O, Auf dem Keller U, Ohler A, Broder C, Schütte A, Kappelhoff R, Stöcker W, Overall CM (2011) Proteomic analyses reveal an acidic prime side specificity for the astacin metalloprotease family reflected by physiological substrates. *Mol Cell Proteomics* 10: M111.009233.
179. Shen Y, Liu T, Tolić N, Petritis BO, Zhao R, Moore RJ, Purvine SO, Camp DG, Smith RD (2010) Strategy for degradomic-peptidomic analysis of human blood plasma. *J Proteome Res* 9: 2339-2346.
180. Shen Y, Tolić N, Liu T, Zhao R, Petritis BO, Gritsenko MA, Camp DG, Moore RJ, Purvine SO, Esteva FJ, Smith RD (2010) Blood peptidome-degradome profile of breast cancer. *PLoS ONE* 5: e13133.
181. Shen Y, Tolić N, Hixson KK, Purvine SO, Anderson GA, Smith RD (2008) De novo sequencing of unique sequence tags for discovery of post-translational modifications of proteins. *Anal Chem* 80: 7742-7754.
182. Zettl M, Adrain C, Strisovsky K, Lastun V, Freeman M (2011) Rhomboid family pseudoproteases use the ER quality control machinery to regulate intercellular signaling. *Cell* 145: 79-91.
183. Adrain C, Zettl M, Christova Y, Taylor N, Freeman M (2012) Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE. *Science* 335: 225-228.
184. Whisstock JC, Silverman GA, Bird PI, Bottomley SP, Kaiserman D, Luke CJ, Pak SC, Reichhart J-M, Huntington JA (2010) Serpins flex their muscle: II. Structural insights into target peptidase recognition, polymerization, and transport functions. *J Biol Chem* 285: 24307-24312.

185. McGraw LA, Gibson G, Clark AG, Wolfner MF (2004) Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster*. *Curr Biol* 14: 1509-1514.
186. Mack PD, Kapelnikov A, Heifetz Y, Bender M (2006) Mating-responsive genes in reproductive tissues of female *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 103: 10358-10363.
187. Lawniczak MKN, Begun DJ (2007) Molecular population genetics of female-expressed mating-induced serine proteases in *Drosophila melanogaster*. *Mol Biol Evol* 24: 1944-1951.
188. McGraw LA, Clark AG, Wolfner MF (2008) Post-mating gene expression profiles of female *Drosophila melanogaster* in response to time and to four male accessory gland proteins. *Genetics* 179: 1395-1408.
189. Peng J, Chen S, Büsler S, Liu H, Honegger T, Kubli E (2005) Gradual release of sperm bound sex-peptide controls female postmating behavior in *Drosophila*. *Curr Biol* 15: 207-213.
190. Prokupek AM, Kachman SD, Ladunga I, Harshman LG (2009) Transcriptional profiling of the sperm storage organs of *Drosophila melanogaster*. *Insect Mol Biol* 18: 465-475.
191. Schnakenberg SL, Matias WR, Siegal ML (2011) Sperm-storage defects and live birth in *Drosophila* females lacking spermathecal secretory cells. *PLoS Biol* 9: e1001192.
192. Mancini E, Tammara F, Baldini F, Via A, Raimondo D, George P, Audisio P, Sharakhov IV, Tramontano A, Catteruccia F, Della Torre A (2011) Molecular evolution of a gene cluster of serine proteases expressed in the *Anopheles gambiae* female reproductive tract. *BMC Evol Biol* 11: 72.
193. Kelleher E, Swanson WJ, Markow TA (2007) Gene duplication and adaptive evolution of digestive proteases in *Drosophila arizonae* female reproductive tracts. *PLoS Genet* 3: e148.
194. Sirot LK, Poulson RL, McKenna MC, Girnary H, Wolfner MF, Harrington LC (2008) Identity and transfer of male reproductive gland proteins of the dengue vector mosquito, *Aedes aegypti*: potential tools for control of female feeding and reproduction. *Insect Biochem Mol Biol* 38: 176-189.

195. Lessley BA, Brown KI (1978) Purification and properties of a proteinase inhibitor from chicken seminal plasma. *Biol Reprod* 19: 223-234.
196. Dean MD, Clark NL, Findlay GD, Karn RC, Yi X, Swanson WJ, Maccoss MJ, Nachman MW (2009) Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (*Mus domesticus*). *Mol Biol Evol* 26: 1733-1743.
197. Chen Y-W, Lee M-S, Lucht A, Chou F-P, Huang W, Havighurst TC, Kim K, Wang J-K, Antalis TM, Johnson MD, Lin C-Y (2010) TMPRSS2, a serine protease expressed in the prostate on the apical surface of luminal epithelial cells and released into semen in prostasomes, is misregulated in prostate cancer cells. *Am J Pathol* 176: 2986-2996.
198. Khan AR, James MN (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci* 7: 815-836.
199. Knowles LL, Markow TA (2001) Sexually antagonistic coevolution of a postmating-prezygotic reproductive character in desert *Drosophila*. *Proc Natl Acad Sci USA* 98: 8692-8696.
200. Ravi Ram K, Wolfner M (2007) Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. *Integrative and Comparative Biology* 47: 427-445.
201. Sirot LK, Hardstone MC, Helinski MEH, Ribeiro JMC, Kimura M, Deewatthanawong P, Wolfner MF, Harrington LC (2011) Towards a semen proteome of the dengue vector mosquito: protein identification and potential functions. *PLoS Negl Trop Dis* 5: e989.
202. Kalb JM, DiBenedetto AJ, Wolfner MF (1993) Probing the function of *Drosophila melanogaster* accessory glands by directed cell ablation. *Proc Natl Acad Sci USA* 90: 8093-8097.
203. Chapman T, Bangham J, Vinti G, Seifried B, Lung O, Wolfner MF, Smith HK, Partridge L (2003) The sex peptide of *Drosophila melanogaster*: female post-mating responses analyzed by using RNA interference. *Proc Natl Acad Sci USA* 100: 9923-9928.
204. Liu H, Kubli E (2003) Sex-peptide is the molecular basis of the sperm effect in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 100: 9929-9933.

205. Bertram MJ, Neubaum DM, Wolfner MF (1996) Localization of the *Drosophila* male accessory gland protein Acp36DE in the mated female suggests a role in sperm storage. *Insect Biochem Mol Biol* 26: 971-980.
206. Adams EM, Wolfner MF (2007) Seminal proteins but not sperm induce morphological changes in the *Drosophila melanogaster* female reproductive tract during sperm storage. *J Insect Physiol* 53: 319-331.
207. Bajaj SP, Rapaport SI, Brown SF (1981) Isolation and characterization of human factor VII. Activation of factor VII by factor Xa. *J Biol Chem* 256: 253-259.
208. Yiallourous I, Kappelhoff R, Schilling O, Wegmann F, Helms MW, Auge A, Brachtendorf G, Berkhoff EG, Beermann B, Hinz HJ, König S, Peter-Katalinic J, Stöcker W (2002) Activation mechanism of pro-astacin: role of the pro-peptide, tryptic and autoproteolytic cleavage and importance of precise amino-terminal processing. *J Mol Biol* 324: 237-246.
209. Amour A, Bird M, Chaudry L, Deadman J, Hayes D, Kay C (2004) General considerations for proteolytic cascades. *Biochem Soc Trans* 32: 15-16.
210. Bloch Qazi MC, Wolfner MF (2006) Emergence of sperm from female storage sites has egg-influenced and egg-independent phases in *Drosophila melanogaster*. *Biol Lett* 2: 128-130.
211. Aguadé M (1998) Different forces drive the evolution of the Acp26Aa and Acp26Ab accessory gland genes in the *Drosophila melanogaster* species complex. *Genetics* 150: 1079-1089.
212. Wong A, Turchin M, Wolfner MF, Aquadro CF (2011) Temporally variable selection on proteolysis related reproductive tract proteins in *Drosophila*. *Mol Biol Evol*.
213. Mattsson JM, Valmu L, Laakkonen P, Stenman U-H, Koistinen H (2008) Structural characterization and anti-angiogenic properties of prostate-specific antigen isoforms in seminal fluid. *Prostate* 68: 945-954.
214. Cui J, Sackton K, Horner VL, Kumar KE, Wolfner MF (2008) Wispy, the *Drosophila* homolog of GLD-2, is required during oogenesis and egg activation. *Genetics* 178: 2017-2029.
215. Xue L, Noll M (2002) Dual role of the Pax gene paired in accessory gland development of



*Drosophila*. Development 129: 339-346.

216. Yamamoto M-T, Takemori N (2010) Proteome profiling reveals tissue-specific protein expression in the male reproductive system of *Drosophila melanogaster*. *Fly* 4: 36-39.
217. Monsma SA, Wolfner MF (1988) Structure and expression of a *Drosophila* male accessory gland gene whose product resembles a peptide pheromone precursor. *Genes Dev* 2: 1063-1073.
218. Monsma SA, Harada HA, Wolfner MF (1990) Synthesis of two *Drosophila* male accessory gland proteins and their fate after transfer to the female during mating. *Dev Biol* 142: 465-475.
219. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, Richardson MF, Anholt RRH, Barrón M, Bess C, Blankenburg KP, Carbone MA, Castellano D, Chaboub L, Duncan L, Harris Z, et al. (2012) The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482: 173-178.
220. Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118: 401-415.
221. Guevara T, Yiallourous I, Kappelhoff R, Bissdorf S, Stöcker W, Gomis-Rüth FX (2010) Proenzyme structure and activation of astacin metallopeptidase. *J Biol Chem* 285: 13958-13965.
222. Tretter V, Altmann F, März L (1991) Peptide-N<sub>4</sub>-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1----3 to the asparagine-linked N-acetylglucosamine residue. *Eur J Biochem* 199: 647-652.
223. Pan Y, Robinett CC, Baker BS (2011) Turning males on: activation of male courtship behavior in *Drosophila melanogaster*. *PLoS ONE* 6: e21144.
224. Gomis-Rüth FX, Stöcker W, Huber R, Zwilling R, Bode W (1993) Refined 1.8 Å X-ray crystal structure of astacin, a zinc-endopeptidase from the crayfish *Astacus astacus* L. Structure determination, refinement, molecular structure and comparison with thermolysin. *J Mol Biol* 229: 945-968.
225. Domínguez A, Albornoz J (1996) Rates of movement of transposable elements in *Drosophila melanogaster*. *Mol Gen Genet* 251: 130-138.

226. Wolfner MF (2009) Battle and ballet: molecular interactions between the sexes in *Drosophila*. *J Hered* 100: 399-410.
227. Hamada FN, Rosenzweig M, Kang K, Pulver SR, Ghezzi A, Jegla TJ, Garrity PA (2008) An internal thermal sensor controlling temperature preference in *Drosophila*. *Nature* 454: 217-220.
228. Manoli DS, Foss M, Villella A, Taylor BJ, Hall JC, Baker BS (2005) Male-specific fruitless specifies the neural substrates of *Drosophila* courtship behaviour. *Nature* 436: 395-400.
229. Pfeiffer BD, Jenett A, Hammonds AS, Ngo T-TB, Misra S, Murphy C, Scully A, Carlson JW, Wan KH, Lavery TR, Mungall C, Svirskas R, Kadonaga JT, Doe CQ, Eisen MB, Celniker SE, Rubin GM (2008) Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proc Natl Acad Sci USA* 105: 9715-9720.
230. Groth AC, Fish M, Nusse R, Calos MP (2004) Construction of transgenic *Drosophila* by using the site-specific integrase from phage  $\phi$ C31. *Genetics* 166: 1775-1782.
231. Bateman JR, Lee AM, Wu C-T (2006) Site-specific transformation of *Drosophila* via  $\phi$ C31 integrase-mediated cassette exchange. *Genetics* 173: 769-777.
232. Dietzl G, Chen D, Schnorrer F, Su K-C, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S, Couto A, Marra V, Keleman K, Dickson BJ (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* 448: 151-156.
233. Markstein M, Pitsouli C, Villalta C, Celniker SE, Perrimon N (2008) Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat Genet* 40: 476-483.
234. Huguenin M, Müller EJ, Trachsel-Rösmann S, Oneda B, Ambort D, Sterchi EE, Lottaz D (2008) The metalloprotease meprin $\beta$  processes E-cadherin and weakens intercellular adhesion. *PLoS ONE* 3: e2153.
235. Serpe M, Ralston A, Blair SS, O'Connor MB (2005) Matching catalytic activity to developmental function: tollid-related processes Sog in order to help specify the posterior crossvein in the *Drosophila* wing. *Development* 132: 2645-2656.
236. Chapman T, Herndon LA, Heifetz Y, Partridge L, Wolfner MF (2001) The Acp26Aa

- seminal fluid protein is a modulator of early egg hatchability in *Drosophila melanogaster*. *Proc Biol Sci* 268: 1647-1654.
237. Lange V, Picotti P, Domon B, Aebersold R (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* 4: 222.
  238. Schmidt A, Claassen M, Aebersold R (2009) Directed mass spectrometry: towards hypothesis-driven proteomics. *Curr Opin Chem Biol* 13: 510-517.
  239. Peña-Castillo L, Hughes TR (2007) Why are there still over 1000 uncharacterized yeast genes? *Genetics* 176: 7-14.
  240. Costello JC, Dalkilic MM, Beason SM, Gehlhausen JR, Patwardhan R, Middha S, Eads BD, Andrews JR (2009) Gene networks in *Drosophila melanogaster*: integrating experimental data to predict gene function. *Genome Biol* 10: R97.
  241. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connolly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, et al. (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285: 901-906.
  242. Flint J, Mackay TFC (2009) Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res* 19: 723-733.
  243. Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 10: 565-577.
  244. Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, Danford T, Bernstein DA, Rolfe PA, Heisler LE, Chin B, Nislow C, Giaever G, Phillips PC, Fink GR, Gifford DK, Boone C (2010) Genotype to phenotype: a complex problem. *Science* 328: 469.
  245. Hrmova M, Fincher GB (2009) Functional genomics and structural biology in the definition of gene function. *Methods Mol Biol* 513: 199-227.
  246. Bréhélin L, Florent I, Gascuel O, Maréchal E (2010) Assessing functional annotation transfers with inter-species conserved coexpression: application to *Plasmodium falciparum*. *BMC Genomics* 11: 35.

247. Walker MG, Volkmut W, Sprinzak E, Hodgson D, Klingler T (1999) Prediction of gene function by genome-scale expression analysis: prostate cancer-associated genes. *Genome Res* 9: 1198-1203.
248. Reverter A, Ingham A, Dalrymple BP (2008) Mining tissue specificity, gene connectivity and disease association to reveal a set of genes that modify the action of disease causing genes. *BioData mining* 1: 8.
249. Vandepoele K, Quimbaya M, Casneuf T, De Veylder L, Van de Peer Y (2009) Unraveling transcriptional control in Arabidopsis using cis-regulatory elements and coexpression networks. *Plant Physiol* 150: 535-546.
250. Klie S, Nikoloski Z, Selbig J (2010) Biological Cluster Evaluation for Gene Function Prediction. *J Comput Biol.*
251. Luo F, Yang Y, Zhong J, Gao H, Khan L, Thompson DK, Zhou J (2007) Constructing gene co-expression networks and predicting functions of unknown genes by random matrix theory. *BMC Bioinformatics* 8: 299.
252. Neubaum DM, Wolfner MF (1999) Wise, winsome, or weird? Mechanisms of sperm storage in female animals. *Curr Top Dev Biol* 41: 67-97.
253. Tram U, Wolfner MF (1999) Male seminal fluid proteins are essential for sperm storage in *Drosophila melanogaster*. *Genetics* 153: 837-844.
254. Carvalho GB, Kapahi P, Anderson DJ, Benzer S (2006) Allogocrine modulation of feeding behavior by the Sex Peptide of *Drosophila*. *Curr Biol* 16: 692-696.
255. Isaac RE, Li C, Leedale AE, Shirras AD (2010) *Drosophila* male sex peptide inhibits siesta sleep and promotes locomotor activity in the post-mated female. *Proc Biol Sci* 277: 65-70.
256. Langfelder P, Horvath S (2007) Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol* 1: 54.
257. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2: 953-971.

258. Saudan P, Hauck K, Soller M, Choffat Y, Ottiger M, Spörri M, Ding Z, Hess D, Gehrig PM, Klauser S, Hunziker P, Kubli E (2002) Ductus ejaculatorius peptide 99B (DUP99B), a novel *Drosophila melanogaster* sex-peptide pheromone. *Eur J Biochem* 269: 989-997.
259. Burn TC, Vigoreaux JO, Tobin SL (1989) Alternative 5C actin transcripts are localized in different patterns during *Drosophila* embryogenesis. *Dev Biol* 131: 345-355.
260. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44-57.
261. Stone EA, Ayroles JF (2009) Modulated modularity clustering as an exploratory tool for functional genomic inference. *PLoS Genet* 5: e1000479.
262. Nüsslein-Volhard C, Wieschaus E (1980) Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287: 795-801.
263. Kilchherr E, Schumaker VN, Phillips ML, Curtiss LK (1986) Activation of the first component of human complement, C1, by monoclonal antibodies directed against different domains of subcomponent C1q. *J Immunol* 137: 255-262.
264. Bertuccioli C, Fasano L, Jun S, Wang S, Sheng G, Desplan C (1996) In vivo requirement for the paired domain and homeodomain of the paired segmentation gene product. *Development* 122: 2673-2685.
265. Xue L, Noll M (1996) The functional conservation of proteins in evolutionary alleles and the dominant role of enhancers in evolution. *EMBO J* 15: 3722-3731.
266. Xue L, Noll M (2000) *Drosophila* female sexual behavior induced by sterile males showing copulation complementation. *Proc Natl Acad Sci USA* 97: 3272-3275.
267. Aravind L (2000) Guilt by association: contextual information in genome analysis. *Genome Res* 10: 1074-1077.
268. Miozzi L, Piro RM, Rosa F, Ala U, Silengo L, Di Cunto F, Provero P (2008) Functional annotation and identification of candidate disease genes by computational analysis of normal tissue gene expression data. *PLoS ONE* 3: e2439.
269. Möhrle F, Baus S, Gruber A, Rackwitz HR, Schnölzer M, Vogt G, Zwillig R (2001)

- Activation of pro-astacin. Immunological and model peptide studies on the processing of immature astacin, a zinc-endoropeptidase from the crayfish *Astacus astacus*. *Eur J Biochem* 268: 2540-2546.
270. Ellison PT, Gray PB, editors (2009) *Endocrinology of social relationships*. Cambridge, MA: Harvard University Press. 499 p.
271. Keverne EB, Curley JP (2004) Vasopressin, oxytocin and social behaviour. *Curr Opin Neurobiol* 14: 777-783.
272. Adkins-Regan E (2009) Hormones and sexual differentiation of avian social behavior. *Dev Neurosci* 31: 342-350.
273. de Bono M, Maricq AV (2005) Neuronal substrates of complex behaviors in *C. elegans*. *Annu Rev Neurosci* 28: 451-501.
274. Smith AA, Hölldobler B, Liebig J (2008) Hydrocarbon signals explain the pattern of worker and egg policing in the ant *Aphaenogaster cockerelli*. *J Chem Ecol* 34: 1275-1282.
275. Villella A, Hall JC (2008) Neurogenetics of courtship and mating in *Drosophila*. *Adv Genet* 62: 67-184.
276. Dickson BJ (2008) Wired for sex: the neurobiology of *Drosophila* mating decisions. *Science* 322: 904-909.
277. Chapman T (2001) Seminal fluid-mediated fitness traits in *Drosophila*. *Heredity* 87: 511-521.
278. Chapman T, Davies SJ (2004) Functions and analysis of the seminal fluid proteins of male *Drosophila melanogaster* fruit flies. *Peptides* 25: 1477-1490.
279. Spieth HT, Ringo JM (1983) Mating behavior and sexual isolation in *Drosophila*. In: Ashburner MM, Carson HL, Thompson JN, editors. *The Genetics and Biology of Drosophila*. New York: Academic Press. pp. 223-284.
280. Ringo J (1996) Sexual receptivity in insects. *Annu Rev Entomol* 41: 473-494.

281. Lawniczak MKN, Begun DJ (2004) A genome-wide analysis of courting and mating responses in *Drosophila melanogaster* females. *Genome* 47: 900-910.
282. Peng J, Zipperlen P, Kubli E (2005) *Drosophila* sex-peptide stimulates female innate immune system after mating via the Toll and Imd pathways. *Curr Biol* 15: 1690-1694.
283. Kapelnikov A, Zelinger E, Gottlieb Y, Rhrissorrakrai K, Gunsalus KC, Heifetz Y (2008) Mating induces an immune response and developmental switch in the *Drosophila* oviduct. *Proc Natl Acad Sci USA* 105: 13912-13917.
284. Fedorka KM, Linder JE, Winterhalter W, Promislow D (2007) Post-mating disparity between potential and realized immune response in *Drosophila melanogaster*. *Proc Biol Sci* 274: 1211-1217.
285. Fowler K, Partridge L (1989) A cost of mating in female fruitflies. *Nature* 338: 760-761.
286. Civetta A, Clark AG (2000) Correlated effects of sperm competition and postmating female mortality. *Proc Natl Acad Sci USA* 97: 13162-13165.
287. Barnes AI, Wigby S, Boone JM, Partridge L, Chapman T (2008) Feeding, fecundity and lifespan in female *Drosophila melanogaster*. *Proc Biol Sci* 275: 1675-1683.
288. Manning A (1962) A sperm factor affecting the receptivity of *Drosophila melanogaster* females. *Nature* 194: 252-253.
289. Manning A (1967) The control of sexual receptivity in female *Drosophila*. *Anim Behav* 15: 239-250.
290. Gilbert D (1981) Ejaculate esterase-6 and initial sperm use by female *Drosophila melanogaster*. *J Insect Physiol* 27: 641-650.
291. Meikle DB, Sheehan KB, Phillis DM, Richmond RC (1990) Localization and longevity of seminal-fluid esterase-t in mated female *Drosophila melanogaster*. *J Insect Physiol* 36: 93-101.
292. Ludwig MZ, Uspensky II, Ivanov AI, Kopantseva MR, Dianov CM, Tamarina NA, Korochkin LI (1991) Genetic control and expression of the major ejaculatory bulb protein (PEB-me) in *Drosophila melanogaster*. *Biochem Genet* 29: 215-239.

293. Samakovlis C, Kylsten P, Kimbrell DA, Engström A, Hultmark D (1991) The andropin gene and its product, a male-specific antibacterial peptide in *Drosophila melanogaster*. *EMBO J* 10: 163-169.
294. Chapman T (2008) The soup in my fly: evolution, form and function of seminal fluid proteins. *PLoS Biol* 6: e179.
295. Pelosi P, Calvello M, Ban L (2005) Diversity of odorant-binding proteins and chemosensory proteins in insects. *Chem Senses* 30 Suppl 1: i291-292.
296. Pilpel N, Nezer I, Applebaum SW, Heifetz Y (2008) Mating-increases trypsin in female *Drosophila* hemolymph. *Insect Biochem Mol Biol* 38: 320-330.
297. Iida K, Cavener DR (2004) Glucose dehydrogenase is required for normal sperm storage and utilization in female *Drosophila melanogaster*. *J Exp Biol* 207: 675-681.
298. Yapici N, Kim Y-J, Ribeiro C, Dickson BJ (2008) A receptor that mediates the post-mating switch in *Drosophila* reproductive behaviour. *Nature* 451: 33-37.
299. Allen AK, Spradling AC (2008) The Sf1-related nuclear hormone receptor Hr39 regulates *Drosophila* female reproductive tract development and function. *Development* 135: 311-321.
300. Prokupek A, Hoffmann F, Eyun S-I, Moriyama E, Zhou M, Harshman L (2008) An evolutionary expressed sequence tag analysis of *Drosophila* spermatheca genes. *Evolution* 62: 2936-2947.
301. McGraw LA, Gibson G, Clark AG, Wolfner MF (2009) Strain-dependent differences in several reproductive traits are not accompanied by early postmating transcriptome changes in female *Drosophila melanogaster*. *Genetics* 181: 1273-1280.
302. Innocenti P, Morrow EH (2009) Immunogenic males: a genome-wide analysis of reproduction and the cost of mating in *Drosophila melanogaster* females. *J Evol Biol* 22: 964-973.
303. Cline TW, Meyer BJ (1996) Vive la différence: males vs females in flies vs worms. *Annu Rev Genet* 30: 637-702.



304. Christiansen AE, Keisman EL, Ahmad SM, Baker BS (2002) Sex comes in from the cold: the integration of sex and pattern. *Trends Genet* 18: 510-516.
305. Manoli DS, Meissner GW, Baker BS (2006) Blueprints for behavior: genetic specification of neural circuitry for innate behaviors. *Trends Neurosci* 29: 444-451.
306. Shirangi TR, McKeown M (2007) Sex in flies: what 'body--mind' dichotomy? *Dev Biol* 306: 10-19.
307. Yamamoto D (2007) The neural and genetic substrates of sexual behavior in *Drosophila*. *Adv Genet* 59: 39-66.
308. Anderson RC (1945) A Study of the Factors Affecting Fertility of Lozenge Females of *Drosophila Melanogaster*. *Genetics* 30: 280-296.
309. Clark AG (2002) Sperm competition and the maintenance of polymorphism. *Heredity* 88: 148-153.
310. Schully SD, Hellberg ME (2006) Positive selection on nucleotide substitutions and indels in accessory gland proteins of the *Drosophila pseudoobscura* subgroup. *J Mol Evol* 62: 793-802.
311. Panhuis TM, Clark NL, Swanson WJ (2006) Rapid evolution of reproductive proteins in abalone and *Drosophila*. *Philos Trans R Soc Lond, B, Biol Sci* 361: 261-268.
312. Swanson WJ, Aquadro CF, Vacquier VD (2001) Polymorphism in abalone fertilization proteins is consistent with the neutral evolution of the egg's receptor for lysin (VERL) and positive darwinian selection of sperm lysin. *Mol Biol Evol* 18: 376-383.
313. Dean MD, Good JM, Nachman MW (2008) Adaptive evolution of proteins secreted during sperm maturation: an analysis of the mouse epididymal transcriptome. *Mol Biol Evol* 25: 383-392.
314. Wong A, Albright SN, Wolfner MF (2006) Evidence for structural constraint on ovulin, a rapidly evolving *Drosophila melanogaster* seminal protein. *Proc Natl Acad Sci USA* 103: 18644-18649.
315. Wong A, Christopher AB, Buehner NA, Wolfner MF (2010) Immortal coils: conserved

- dimerization motifs of the *Drosophila* ovulation prohormone ovulin. *Insect Biochem Mol Biol* 40: 303-310.
316. Aguadé M, Miyashita N, Langley CH (1992) Polymorphism and divergence in the Mst26A male accessory gland gene region in *Drosophila*. *Genetics* 132: 755-770.
317. Tsaur SC, Wu CI (1997) Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of *Drosophila*. *Mol Biol Evol* 14: 544-549.
318. Tsaur SC, Ting CT, Wu CI (1998) Positive selection driving the evolution of a gene of male reproduction, Acp26Aa, of *Drosophila*: II. Divergence versus polymorphism. *Mol Biol Evol* 15: 1040-1046.
319. Kern AD, Jones CD, Begun DJ (2004) Molecular population genetics of male accessory gland proteins in the *Drosophila simulans* complex. *Genetics* 167: 725-735.
320. Andolfatto P (2005) Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437: 1149-1152.
321. Tamura K, Subramanian S, Kumar S (2004) Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol* 21: 36-44.
322. Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155: 1405-1413.
323. Clark AG, Aguadé M, Prout T, Harshman LG, Langley CH (1995) Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. *Genetics* 139: 189-201.
324. Clark AG, Dermitzakis ET, Civetta A (2000) Nontransitivity of sperm precedence in *Drosophila*. *Evolution* 54: 1030-1035.
325. Fiumera AC, Dumont BL, Clark AG (2005) Sperm competitive ability in *Drosophila melanogaster* associated with variation in male reproductive proteins. *Genetics* 169: 243-257.
326. Clark AG, Begun DJ (1998) Female genotypes affect sperm displacement in *Drosophila*. *Genetics* 149: 1487-1493.

327. Clark AG, Begun DJ, Prout T (1999) Female x male interactions in *Drosophila* sperm competition. *Science* 283: 217-220.
328. Civetta A, Clark AG (2000) Chromosomal effects on male and female components of sperm precedence in *Drosophila*. *Genet Res* 75: 143-151.
329. Bjork A, Starmer WT, Higginson DM, Rhodes CJ, Pitnick S (2007) Complex interactions with females and rival males limit the evolution of sperm offence and defence. *Proc Biol Sci* 274: 1779-1788.
330. Chippindale AK, Gibson JR, Rice WR (2001) Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila*. *Proc Natl Acad Sci USA* 98: 1671-1675.
331. Rice WR (2000) Dangerous liaisons. *Proc Natl Acad Sci USA* 97: 12953-12955.
332. Gavrillets S, Arnqvist G, Friberg U (2001) The evolution of female mate choice by sexual conflict. *Proc Biol Sci* 268: 531-539.
333. Civetta A (2003) Shall we dance or shall we fight? Using DNA sequence data to untangle controversies surrounding sexual selection. *Genome* 46: 925-929.
334. Lawniczak MKN, Begun DJ (2005) A QTL analysis of female variation contributing to refractoriness and sperm competition in *Drosophila melanogaster*. *Genet Res* 86: 107-114.
335. Holland B, Rice WR (1999) Experimental removal of sexual selection reverses intersexual antagonistic coevolution and removes a reproductive load. *Proc Natl Acad Sci USA* 96: 5083-5088.
336. Crudgington HS, Beckerman AP, Brüstle L, Green K, Snook RR (2005) Experimental removal and elevation of sexual selection: does sexual selection generate manipulative males and resistant females? *Am Nat* 165 Suppl 5: S72-87.
337. Rice WR (1996) Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature* 381: 232-234.
338. Wigby S, Sirot LK, Linklater JR, Buehner N, Calboli FCF, Bretman A, Wolfner MF,

- Chapman T (2009) Seminal fluid protein allocation and male reproductive success. *Curr Biol* 19: 751-757.
339. Bretman A, Fricke C, Chapman T (2009) Plastic responses of male *Drosophila melanogaster* to the level of sperm competition increase male reproductive fitness. *Proc Biol Sci* 276: 1705-1711.
340. Swanson WJ, Wong A, Wolfner MF, Aquadro C (2004) Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics* 168: 1457-1465.
341. Clark NL, Gasper J, Sekino M, Springer SA, Aquadro CF, Swanson WJ (2009) Coevolution of interacting fertilization proteins. *PLoS Genet* 5: e1000570.
342. Swanson WJ, Vacquier VD (1998) Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science* 281: 710-712.
343. McGraw LA, Fiumera AC, Ramakrishnan M, Madhavarapu S, Clark AG, Wolfner MF (2007) Larval rearing environment affects several post-copulatory traits in *Drosophila melanogaster*. *Biol Lett* 3: 607-610.
344. Fricke C, Bretman A, Chapman T (2008) Adult male nutrition and reproductive success in *Drosophila melanogaster*. *Evolution* 62: 3170-3177.
345. Prout T, Clark AG (1996) Polymorphism in genes that influence sperm displacement. *Genetics* 144: 401-408.
346. Fiumera AC, Dumont BL, Clark AG (2006) Natural variation in male-induced 'cost-of-mating' and allele-specific association with male reproductive genes in *Drosophila melanogaster*. *Philos Trans R Soc Lond, B, Biol Sci* 361: 355-361.
347. Mackay T, Richards S, Gibbs R (2008) Proposal to Sequence a *Drosophila* Genetic Reference Panel: A Community Resource for the Study of Genotypic and Phenotypic Variation.
348. Heifetz Y, Tram U, Wolfner MF (2001) Male contributions to egg production: the role of accessory gland products and sperm in *Drosophila melanogaster*. *Proc Biol Sci* 268: 175-180.

349. Jagadeeshan S, Singh RS (2007) Rapid evolution of outer egg membrane proteins in the *Drosophila melanogaster* subgroup: a case of ecologically driven evolution of female reproductive traits. *Mol Biol Evol* 24: 929-938.
350. Buszczak M, Freeman MR, Carlson JR, Bender M, Cooley L, Segraves WA (1999) Ecdysone response genes govern egg chamber development during mid-oogenesis in *Drosophila*. *Development* 126: 4581-4589.
351. Pascucci T, Perrino J, Mahowald AP, Waring GL (1996) Eggshell assembly in *Drosophila*: processing and localization of vitelline membrane and chorion proteins. *Dev Biol* 177: 590-598.
352. Tzolovsky G, Deng WM, Schlitt T, Bownes M (1999) The function of the broad-complex during *Drosophila melanogaster* oogenesis. *Genetics* 153: 1371-1383.
353. Pierceall WE, Li C, Biran A, Miura K, Raikhel AS, Segraves WA (1999) E75 expression in mosquito ovary and fat body suggests reiterative use of ecdysone-regulated hierarchies in development and reproduction. *Mol Cell Endocrinol* 150: 73-89.
354. Carney GE, Bender M (2000) The *Drosophila* ecdysone receptor (EcR) gene is required maternally for normal oogenesis. *Genetics* 154: 1203-1211.
355. Kim C, Han K, Kim J, Yi JS, Kim C, Yim J, Kim Y-J, Kim-Ha J (2002) Femcoat, a novel eggshell protein in *Drosophila*: functional analysis by double stranded RNA interference. *Mech Dev* 110: 61-70.
356. Tootle TL, Spradling AC (2008) *Drosophila* Pxt: a cyclooxygenase-like facilitator of follicle maturation. *Development* 135: 839-847.
357. Tootle TL, Williams D, Hubb A, Frederick R, Spradling A (2011) *Drosophila* eggshell production: identification of new genes and coordination by Pxt. *PLoS ONE* 6: e19943.
358. Cavaliere V, Spanò S, Andrenacci D, Cortesi L, Gargiulo G (1997) Regulatory elements in the promoter of the vitelline membrane gene VM32E of *Drosophila melanogaster* direct gene expression in distinct domains of the follicular epithelium. *Mol Gen Genet* 254: 231-237.